METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT

DIRETTORE PROPRIETARIO — DIRECTEUR ET PROPRIÉTAIRE
EDITOR AND PROPRIETOR — HERAUSGEBER UND EIGENTHÜMER

Prof. Dott. Corrado Gini, *Direttore dell'Istituto di Statistica della R. Università di Roma.*

COMITATO DIRETTIVO — COMITÉ DE DIRECTION
EDITORIAL COMMITTEE — DIREKTION-KOMITEE

Prof. A. Andréadès, *prof. de Science des finances à l'Université d'Athènes* (Grèce).
Prof. F. Bernstein, *früherer Direktor des Instituts für mathematische Statistik der Universität, Göttingen* (Deutschland), *jetzt an der Columbia University* (U. S. A).
Prof. A. E. Bunge, *director gen. de Estadística de la Nación, Buenos Aires* (Argentina).
Prof. F. P. Cantelli, *professore di Matematica Attuariale nel R. Istituto Superiore di Scienze Economiche e Commerciali di Roma* (Italia).
Prof. C. V. L. Charlier, *professor der Astronomie an der Universität Lund* (Schweden).
Prof. A. Flores de Lemus, *jefe de Estadística del Min. de Hacienda, Madrid* (España).
Prof. M. Greenwood, *professor of Epidemiology and Vital Statistics in the University of London* (England).
Dott. G. Jahn, *directeur du Bureau Central de Statistique de Norvège, Oslo* (Norvège).
Prof. A. Julin, *secrétaire général honoraire du Ministère de l'Industrie, du Travail et de la Prévoyance sociale, Bruxelles* (Belgique).
Prof. H. W. Methorst, *directeur de l'Office permanent de l'Institut International de Statistique et du Bureau central de Statistique, La Haye* (Pays-Bas).
Prof. W. F. Ogburn, *professor of Sociology in the University of Chicago* (U. S. A.).
Prof. R. Pearl, *director of the Department of Biology of the School of Hygiene and Public Health, Baltimore* (U. S. A.).
Prof. H. Westergaard, *professor in the University of Copenhagen* (Denmark).

AMMINISTRATORE — ADMINISTRATEUR — MANAGER — VERWALTER
Dott. Silvio Orlandi, *Istituto di Statistica della R. Università di Roma.*

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION
EDITORIAL SECRETARIES — REDACTIONSSECRETAERE
Prof. Luigi Galvani — Prof. Mario Saibante

Vol. XII - N. 1 **31-XII-1934.**

SOMMARIO — SOMMAIRE — CONTENTS — INHALT

ROMA
AMMINISTRAZIONE DEL «METRON»
R. UNIVERSITÀ — ISTITUTO DI STATISTICA

ARTICOLI GIUNTI ALLA RIVISTA CHE VERRANNO PUBBLICATI NEI PROSSIMI NUMERI.

(Secondo l'ordine d'arrivo).

ARTIKEL DIE AN DIE ZEITSCHRIFT ANGELANGT SIND UND WELCHE IN DEN NACHFOLGENDEN NUMMERN ERSCHEINEN WERDEN.

(Nach der Reihenfolge des Eingangs)

ARTICLES REÇUS PAR LA REVUE ET À PARAÎTRE PROCHAINEMENT.

(D'après la date de reception).

ARTICLES RECEIVED BY THE REVIEW WHICH WILL BE PUBLISHEED IN FUTURE ISSUES.

(According to date of receipt)

N. Smirnoff. *Ueber die Verteilung des allgemeinen Gliedes in der Variationsreihe.*

J. O. Irwin. *Tests of Significance for differences between percentages based on small numbers.*

H. Koeppler. *Das Fehlergesetz des Korrelationskoeffizienten und andere Wahrscheinlichkeitsgesetze der Korrelationstheorie.*

S. Koller. *Die Analyse der Abhängigkeitsverhältnisse in zwei Korrelationssystemen.*

C. E. Dieulefait. *Généralisation des courbes de K. Pearson.*

E. Zwinggi. *Zur Frage des Beharrungszustandes.*

H. Wold. *A Study on the Mean Difference, Concentration Curves, and the Report of Concentration.*

L. I. Dublin, A. J. Lotka a **M. Spiegelman.** *Construction of Life Table by Correlation.*

R. Mogno. *Di un metodo di interpolazione statistica.*

S. Kullback. *A Note on the Distribution of a certain partial belonging coefficient.*

Gli Autori degli articoli inviati per la pubblicazione nella Rivista, rinunciano in favore della medesima alla proprietà letteraria degli articoli stessi, qualora vengano pubblicati.

The Authors of papers sent for publication in the Review are supposed to give up their copyright in favour of the Review if the papers are published.

Les Auteurs des articles envoyés à la Revue, pour y être publiés, renoncent, en faveur de celle-ci, à la propriéte littéraire de leurs articles, s'ils sont acceptés.

Die Verfasser der zur Veröffentlichung in der Zeitschrift zugesandten Aufsätzs, werden, falls selbige veröffentlicht werden, auf ihre Verfasserrechte zu Gunsten der Zeitschrift verzichten müssen.

LIEBMANN HERSCH

Essai sur les variations périodiques et leur mensuration

TABLE DES MATIÈRES

Variation comparée des coefficients de la concentration et du nivellement de séries hétéronomes (§ 53). Les limites des variations des indices de la précipitation (§ 54). Coefficients de la précipitation et de la gradation des variations dans le cas de séries ordinaires et dans celui de séries périodiques (§§ 55 et 56). Limites des variations de l'indice K (§ 57). Le coefficient de la concentration k_n (§ 58). Mensuration de la concentration basée sur les carrés des écarts du premier ordre et celle basée sur les écarts du second ordre : les indices C_n et K (§ 59) ; conditions dans lesquelles le coefficient k_n varie en ligne droite (§ 60) ; k_n est moins exact que c_n (§ 61), mais plus sensible aux variations des petites concentrations (§ 62).

CHAPITRE IX. — *Coefficients de certains caractères des séries périodiques* (suite).

Necessité de l'emploi simultané de l'indice de dissymétrie positive et de celui de dissymétrie négative (§ 63). Le coefficient de la symétrie s (§ 64). Coefficient de symétrie s et coefficient de corrélation r (§ 65). Formules du coefficient s (§ 66). Son application aux séries ordinaires et aux séries périodiques (§ 67). Illustrations sur des exemples concrets ; le coefficient de symétrie s et les indices de dissymétrie D (§ 68).

CHAPITRE X. — *Applications à l'étude de quelques périodicités réelles.*

Système de caractères à relever (§ 69). Périodicité annuelle de la température de l'air atmosphérique et de celle du sol à Berlin : tableau des caractères de ces périodicités (§ 70), analyse du tableau (§ 71) ; indices et coefficients (§ 72) ; représentation par des sinusoïdes (§ 73 et 74). Périodicité annuelle de la mortalité infantile en Suisse, en France et dans l'U. R. S. S. (§ 75). Périodicité annuelle du chômage dans l'industrie du bâtiment au Royaume-Uni et au Danemark (§ 76). Périodicités hebdomadaires et annuelle de la consommation du gaz d'éclairage à Genève (§ 77).

ANNEXES :

Annexe I. Tables des coefficients pour séries à 12, à 24 et à 7 termes :

I, A : coefficients de concentration (c) en fonction de $\frac{\sigma}{d_m}$ et coefficients de précipitation (p) des séries périodiques monocéphales en fonction de $\frac{\varsigma}{e_m}$. — *I, B* : *idem* en fonction respectivement de $\frac{\Sigma d^2}{\Sigma^2 d}$ et de $\frac{\Sigma e^2}{4\,\alpha^2}$. — *II, A* : coefficients de nivellement (ν) en fonction de $\frac{d_m}{\sigma}$ et cofficients de gradation (g) des séries périodiques monocéphales en fonction de $\frac{e_m}{\varsigma}$. — *II, B* : *idem* en fonction respectivement de $\frac{\Sigma^2 d}{\Sigma d^2}$ et de $\frac{4\,\alpha^2}{\Sigma e^2}$.

Annexe II. Tableaux statistiques.

PREFACE

Par le travail que nous nous permettons de présenter ici à l'attention du lecteur, nous voulons nous acquitter d'une dette que nous avons contractée il y a plus de quatre ans.

En effet, nous avons déjà annoncé la présente recherche dans la *Revue Internationale du Travail* du mois de janvier 1929, dans la première partie de notre étude intitulée : *Les fluctuations saisonnières du chômage dans l'industrie du bâtiment de certains pays européens* (*). De plus, dans l'étude sur le chômage saisonnier dans l'industrie du bâtiment, nous avons employé certaines methodes basées sur des propositions dont la démonstration n'était pas donnée, mais était expressément renvoyée au présent travail. Tel fut notamment le cas des coefficients de la concentration et du nivellement des variations périodiques (que nous avons appelés alors « indices du degré de concentration » et « de prolongation »). Nous l'avons fait dans l'idée que l'étude methodologique sur *les variations périodiques et leur mensuration* allait paraître à peu près simultanément avec l'étude des faits relatifs au chômage saisonnier.

Cet espoir ne s'est pourtant pas réalisé. Diverses circonstances, étrangères pour la plupart au présent travail, ont beaucoup retardé sa rédaction définitive. Mais ayant repris les mêmes problèmes, en vue de leur mise au point, après une interruption de quelques années, nous avons été amené à modifier, à préciser ou à compléter sur bien des points nos déductions primitives. Nous voudrions donc croire que, paraissant plus tard que nous ne l'avons supposé, cette recherche présente au moins le fruit de réflexions plus mûres.

Entre temps, la notion de la diversité des variations périodiques a gagné beaucoup de terrain dans les conception statistiques. Il y a encore quelques années, rares étaient les statisticiens qui soupçonnaient l'existence, dans la vie sociale, de périodicités plus longues que les fameux « cycles économiques ». Combien nous en sommes

---

(*) Parue aussi comme tirage à part, édition du « Bureau International du Travail », Genève 1929.

loin aujourd'hui après les travaux de L. March, de Simiand et de Wagemann qui ont rendu courante l'idée de cycles semi-séculaires dans la vie économique, après les investigations de Gini qui remettent en honneur l'idée (qui réapparaît elle aussi en quelque sorte périodiquement) d'après laquelle les nations elles-mêmes formeraient au fond des cycles de très longue durée dans l'évolution générale de l'espèce humaine, — pour ne mentionner que ces travaux-là. Une étude sur la nature des variations périodiques et les méthodes de leur mensuration nous paraît aujourd'hui plus actuelle que jamais.

D'un autre côté, nous voudrions dès maintenant relever le fait qu'au cours de notre recherche sur les variations périodiques, nous avons été obligé d'examiner certains problèmes qui peuvent se poser au sujet de toute variation de phénomènes réels, qu'elle soit périodique ou non. En d'autres termes, à l'occasion d'un phénomène relativement particulier, nous étions souvent obligé de nous poser des problèmes d'un ordre plus général. Dans ces conditions, les solutions auxquelles nous sommes arrivé sont parfois telles qu'elles ont une *portée plus générale*, valant aussi bien pour les variations non-périodiques que pour les variations périodiques. Tel est notamment le cas de nos coefficients de concentration et de nivellement, de nos *indices* de précipitation et de gradation ainsi que de nos coefficients de symétrie. Dans d'autres cas, pour pouvoir être étendues aux variations non-périodiques, les solutions adoptées dans cet *Essai* demandent encore une rectification appropriée (tel est le cas des *coefficients* de précipitation et de gradation). Dans certains cas, la solution adoptée ici est pour des séries non-périodiques d'une application plus simple encore que pour les variations périodiques ; tels nos coefficients et indices de symétrie.

En bonne logique, l'examen des problèmes d'ordre plus général aurait dû être détaché du présent travail pour faire l'objet d'une recherche indépendante, précédant l'étude sur les variations périodiques. Pour raison d'opportunité, nous traitons cependant ces problèmes avec les autres (relatifs aux variations proprement périodiques), comme ils se sont posés à nous-même au cours de la présente recherche. Peut-être tel lecteur se montrera-t-il plus indulgent pour les défauts et les lacunes de cette contribution à l'étude de la périodicité des phénomènes si sous certains autres rapports ce même travail apporte plus que son titre ne le fait supposer.

Un mot encore. Il est possible que tel point traité ici ait été déjà effleuré ou même étudié à fond par un autre chercheur sans que

l'auteur de ces lignes le sache. Si tel était le cas, je prierais la critique ou l'auteur lésé lui-même de bien vouloir me signaler ma faute, soit par la presse soit par correspondance et, de mon côté, je me ferais un plaisir autant qu'un devoir de la corriger publiquement à la première occasion.

L. Hersch.

Université de Genève, le 4 mars 1933.

CHAPITRE I.

Observations préliminaires et premières distinctions entre séries de variations périodiques.

§ 1. — Depuis les travaux de Kemmerer et de Persons, on s'occupe beaucoup des méthodes à appliquer pour fixer aussi strictement que possible les fluctuations périodiques des phénomènes. Mais jusqu'à ces dernières années, ces recherches méthodologiques étaient limitées à des problèmes d'un genre spécial. D'un côté, il s'agissait pour ainsi dire toujours de fluctuations *saisonnières* ou, plus exactement, de fluctuations selon les mois de l'année, donc de séries dans lesquelles la période se compose de 12 termes. D'un autre côté, on avait presque toujours en vue des variations de phénomènes d'ordre économique. En outre, on cherchait le plus souvent les méthodes le mieux appropriées pour fixer la « saisonnalité » des phénomènes non pas pour elle-même, mais au contraire pour son *élimination*, afin de mieux déterminer l'état de la conjoncture économique indépendamment des influences passagères de la saison.

En effet, dans la situation momentanée du marché on distinguait habituellement l'effet combiné de quatre facteurs : 1°) le cycle économique, soit la conjoncture proprement dite ; 2°) le *trend*, ou la tendance générale de l'évolution vers la hausse ou la baisse du phénomène observé (prix, production, etc.) ; 3°) la saison et 4°) l'ensemble des circonstances extraordinaires, fortuites (guerres, troubles, catastrophes naturelles, etc.) qui peuvent influencer la marche des phénomènes. On cherchait donc à éliminer la part attribuable aux trois derniers facteurs (trend, saison, hasard) afin de mettre en évidence ce qui caractérise proprement le premier, la conjoncture économique.

Notons tout de suite que si la prise en considération des quatre facteurs énumérés seuls se justifie par la nécessité de simplifier un problème qui est déjà assez compliqué, elle n'en reste pas moins conventionnelle. Car dans la réalité le nombre des facteurs systématiques est bien plus considérable, la succession annuelle des saisons n'étant nullement le seul facteur périodique qui influe sur les phénomènes économiques (indépendamment de la « période » la plus capricieuse, la moins périodique, qu'on appelle le cycle économique). Pour ne pas chercher trop loin, rappelons seulement que les valeurs de bourse par exemple subissent généralement le contre-coup des liquidations de fin de mois (« ultima ») et que les cours cotés alors diffèrent systématiquement de ceux observés à d'autres dates : les observations journalières ou hebdomadaires de certains faits économiques marquent ainsi non seulement une périodicité annuelle, où les 12 mois de l'année constituent une période qui a ses moments caractéristiques (saisons) de hausse et de baisse, mais encore une périodicité mensuelle où les 30 jours forment également une période ayant son maximum et son minimum propre. On peut éliminer l'action de cette périodicité mensuelle (en prenant par exemple la moyenne du mois ou seulement le chiffre observé à une certaine date du mois, par exemple toujours le 15 ou toujours le dernier jour du mois, etc.), tout comme on cherche à éliminer l'influence des autres facteurs, mais la périodicité mensuelle n'en existe pas moins.

§ 2. — D'ailleurs, tout le monde sait que les mouvements périodiques ont, dans la vie économique et sociale comme dans la vie organique et dans le monde inorganique de notre planète (laissant même de côté les périodicités les plus strictes et les plus variées du domaine astronomique), des durées fort variées. Nous connaissons tous des périodes hebdomadaires et des périodes journalières. En effet, le mouvement de voyageurs dans les tramways urbains, le débit des grands magasins, des coopératives, des restaurants, la consommation du gaz ou de l'électricité dans les agglomérations urbaines, les nombres des accidents de circulation ou autres, les nombres des crimes, des suicides, etc., etc. ne se distribuent pas indifféremment entre les divers jours de la semaine ni entre les 24 heures de la journée, comme ils ne se répartissent pas indifféremment entre les 12 mois de l'année, ou tout comme les températures de l'air, du sol, de l'eau, du corps humain, ou le nombre des décès ne se distribuent pas simplement au hasard entre les mêmes 24 heures de la journée. On peut ainsi constater, dans les variations des phénomènes, des périodes plus ou moins

rigoureuses de 24 heures, d'une semaine, d'un mois, d'une année, d'un certain nombre d'années (7-8 environ, qui paraît le mieux caractériser la durée du cycle économique). Et rien ne prouve que ces cinq espèces de prériodicités soient les seules qui existent dans la vie sociale. Même a priori on pourrait en douter beaucoup.

D'ailleurs, qu'il existe dans la société humaine des périodes moindres que de 24 heures, on peut le voir continuellement par exemple dans la vie scolaire où les successions de travail et de récréation forment un mouvement périodique, la période ayant généralement la durée d'une heure.

D'un autre côté, tout porte à supposer qu'il doit y avoir aussi des périodes dont la durée est bien plus considérable que celle du cycle économique. Seulement, l'observation statistique de la vie sociale n'est pas encore suffisamment âgée pour avoir pu relever positivement une série de chiffres comprenant au moins trois périodes dont chacune aurait une durée dépassant celle du cycle économique à peu près dans la même proportion (7-8 fois) que celle-ci dépasse la durée d'une période annuelle. En présence de certaines données historiques, on a pu cependant se demander par exemple s'il n'existe pas dans la vie sociale une périodicité dont les phases correspondent à la succession des générations, chaque phase comprenant environ 25 à 30 ans.

Et en effet, depuis quelques années on est de plus en plus enclin à reconnaître, notamment dans le mouvement des prix, l'existence de cycles « semi-séculaires ». Ces cycles semi-séculaires ne peuvent naturellement pas être limités au seul mouvement des prix ; ils sont nécessairement liés à l'ensemble de la vie sociale et affectent à leur tour la marche de celle-ci. Quant à nous-même, nous avons toujours soutenu l'hypothèse de l'existence de cycles « semi-séculaires » (d'une durée de 50 à 60 ans) non seulement dans le mouvement des prix, mais aussi dans l'ensemble de l'évolution sociale (*).

(*) Ainsi, dès notre premier cours de statistique générale fait à l'Université de Genève en hiver 1915-1916, nous n'avons cessé de souligner qu'une périodicité « semi-séculaire » dont chaque phase comprend de 25 à 30 ans existe très probablement dans l'ensemble de l'histoire, même dans les évènements d'apparence les plus éloignés du mouvement des prix, tels que la succession d'époques de paix et d'époques de guerres et de bouleversements sociaux. Pour ne pas trop nous éloigner dans le passé, nous rappelions qu'après la cessation de la guerre dite de sept ans (1763), l'Europe avait connu une phase de paix relative de plus de 25 ans, jusqu'à la Révolution française

En même temps, nous trouvions a priori fort peu probable que les périodicités dans les phénomènes sociaux soient limités aux quelques espèces énumérées ; nous admettions, au contraire, comme infiniment plus probable qu'il devait exister des périodicités en nombre indéfini, dont les unes pourraient avoir une durée inférieure même à une heure et dont les autres auraient une durée dépassant de beaucoup le demi-siècle. Après tant d'enseignements de l'histoire, nous

---

(1789) ; vint ensuite une phase de guerres et de bouleversements intérieurs qui a duré elle aussi plus de 25 ans, jusqu'après la conclusion du traité de Vienne (1815-18) ; une nouvelle phase de paix et de stabilité intervient alors qui dure jusque vers 1847, c'est-à-dire une trentaine d'années ; ce fut ensuite l'avènement d'une nouvelle phase de bouleversements intérieurs et de guerres qui établirent le régime constitutionnel dans l'Europe centrale et occidentale, qui affranchirent les serfs en Russie, qui créèrent les unités allemande et italienne, qui fondèrent la Hongrie et délivrèrent les Slaves balkaniques, phase qui se termina avec la guerre russo-turque (1877-78), c'est-à-dire de nouveau après une durée de 30 ans ; après, ce fut de nouveau un temps de paix et de stabilité relatives qui dura environ 25 ans, jusqu'à la guerre russo-japonaise (1903-1904) ; à ce moment commença une nouvelle phase de guerres et de révolutions qui est marquée par la première révolution russe, les révolutions en Perse, en Chine, en Turquie, les guerres italo-turque et balkaniques, la guerre mondiale, la nouvelle révolution russe et l'avènement du bolchévisme, les révolutions dans les pays de l'Europe centrale et l'établissement du régime républicain dans ces pays avec accès de leurs partis ouvriers au pouvoir, la résurrection de la Pologne et la création de nouveaux Etats en Europe et en dehors d'elle, la création de la Société des Nations, l'avènement du fascisme et de la dictature en Italie et dans certains autres pays, la révolution en Espagne, les guerres greco-turques, russo-polonaise, marocaine, la guerre civile et la transformation politique de la Chine, l'agitation aux Indes, la campagne japonaise en Maudchourie, etc., etc.

(On remarquera que si cette hypothèse était vraie — et il ne s'agit naturellement que d'une simple hypothèse — nous nous trouverions actuellement enfin sur le seuil d'une phase de paix et de stabilité, et l'avènement d'une nouvelle phase troublée serait à prévoir pour 1960 environ. Malheureusement, l'état d'insécurité est aujourd'hui tel que l'avènement de la paix internationale et de la stabilité intérieure paraît aujourd'hui moins probable qu'une quinzaine d'années encore de guerres et de troubles ne semblait possible en 1916).

Nous rappelions en même temps que ces phases de paix et de guerre (ou de stabilité et d'instabilité sociale) coïncidaient dans une large mesure avec les époques de bon marché et de cherté qu'accuse l'histoire des prix, à savoir que la phase de cherté coïncidait notablement avec celle des guerres et des troubles et la phase de bon marché avec celle de paix et de stabilité (La chute des prix qui se poursuit depuis plusieurs années, particulièrement depuis le début de la crise, deviendrait ainsi d'autant plus significative).

croyions trouver quelques indices de l'existence de cycles *multi-séculaires* notamment dans l'évolution démographique récente des pays occidentaux, évolution qui annonce la fin du mouvement ascendant et l'avènement d'une phase descendante dans le développement ultérieur de leurs populations.

Récemment, CORRADO GINI a développé une théorie d'après laquelle l'histoire des nations présenterait un mouvement cyclique de hausse et de baisse de la population (*). Sur le terrain plus proprement démographique, il se rattache ainsi au courant d'idées qui, sur d'autres terrains, a été représenté au XVIII[e] siècle par la *Science Nouvelle* de VICO et au XIX[e] par l'*Histoire du développement intellectuel de l'Europe* de DRAPER, ainsi qu'aux vues recemment développées par M. FLINDERS PETRIE (**), d'après lequel les périodes dans la vie de l'humanité sont marquees chacune par une certaine civilisation, par une « grande année » des Etrusques ; pour les civilisations méditerranéennes (auxquelles appartiennent celles de l'Egypte, de la Grèce, de Rome et celles de l'Europe moderne), pour celles du Proche Orient (les civilisations assyro-babylonienne, persane, étrusque, hébraïque et arabe) de même que pour celles des Indes et du Mexique, ces périodes peuvent, d'après cet auteur, non seulement être constatées, mais aussi mesurées, et la durée de telles périodes serait à peu près la même dans les diverses parties du monde, s'étendant à environ 15 siècles, dont huit ou neuf constitueraient la phase de hausse et six la phase de baisse.

Dans cet ordre d'idées, on se rappellera aussi les époques « critiques » et les époques « organiques » que SAINT-SIMON a distinguées dans l'histoire de l'humanité sans pourtant attribuer à ce mouvement de flux et de reflux le caractère d'une périodicité proprement dite.

D'ailleurs, il se peut qu'à côté des cycles à durée constante, il en existe aussi d'autres dont la durée varie selon des lois déterminées (***).

(*) *The Cyclical Rise and Fall of Population* dans la série des conférences *Population* publiée par *The University of Chicago Press*, Chicago 1930. Une édition italienne des conférences du Prof. GINI a paru avec quelques additions sous le titre *Nascita, evoluzione e morte delle Nazioni* (Istituto Nazionale Fascista di Cultura, Roma, 1930). Pour plus de détails, cf. l'ouvrage plus recent du même auteur *Le basi scientifiche della politica della popolazione*, Catania, S. E. M., 1931.

(**) *The Revolutions of Civilization* (3d ed., London, Harper 1922).

(***) La prolongation récente de la durée de la vie moyenne dans les pays de la civilisation occidentale ne pourrait-elle pas avoir pour effet l'accroissement de la durée des cycles « semi-séculaires » (qui nous semblent correspondre à la succession des générations) ?

Cette durée peut aller en augmentant ou en diminuant ou encore suivre elle-même un mouvement périodique d'une certaine espèce.

On dira que par ces suppositions nous quittons le terrain positif. Peut-être. Mais précisément en perdant de vue ces diverses possibilités et en projetant dans un avenir quelque peu reculé, des séries forcément restreintes d'observations passées, certains statisticiens — et pas des moindres — sont souvent arrivés à des conclusions qui furent plus tard cruellement démenties par la réalité.

Il se peut en particulier que ce qu'on considère comme le *trend* des phénomènes (et que l'on représente fréquemment encore comme une évolution en ligne droite) forme souvent lui-même une des phases de quelque mouvement périodique de durée plus longue que celle du cycle économique ou même du cycle semi-séculaire.

Toutes ces observations ne visent nullement à contester l'utilité pratique du postulat des quatre facteurs pour la fixation de la conjoncture économique. Elles soulignent seulement le fait que ne voyant, en dehors du cycle économique, qu'une seule espèce de fluctuations périodiques, les fluctuations saisonnières, ce postulat conventionnel, et sans doute inexact, laisse dans l'ombre la grande variété réelle de la périodicité des phénomènes.

§ 3. — Inutile de dire que les variations périodiques peuvent être étudiées non seulement en vue de leur élimination, mais aussi, et peut-être surtout, pour elles-mêmes, tout comme les autres caractères des phénomènes. Ces variations, ne constituent-elles pas un des aspects de la réalité ? (*). Il y a donc un intérêt théorique et pratique à les connaître. Le physiologiste par exemple peut observer la tempé-

---

(*) Il est permis de croire — et la « raison pratique » nous commande même de croire — que toute la vie sociale de l'humanité présente une combinaison de mouvements « évolutifs » et de fluctuations périodiques. Cette conception, connue encore des Arabes au moyen-âge, développée par VICO et rendue populaire par la célèbre image de la spirale de GOETHE, est aujourd'hui pour ainsi dire universellement admise. Mais la raison pratique mise de côté, elle ne paraît pas s'imposer plus que l'hypothèse selon laquelle il n'y aurait en réalité que des combinaisons de divers mouvements périodiques ; en tout cas, dans le domaine de la réalité qui est relativement le moins complexe et où la nature des phénomènes est par suite le mieux observable, en astronomie, on n'admet, je crois, que des combinaisons de mouvements rotatoires, c'est-à-dire périodiques. Quoi qu'il en soit, l'existence de mouvements périodiques, me semble-t-il, peut *le moins* être mise en doute.

rature du corps humain à diverses heures de la journée non seulement en vue d'établir sa température moyenne (élimination des variations d'une heure à l'autre), mais aussi dans le but de connaître la marche et l'étendue de ces variations horaires, de les comparer chez des individus d'âge et de sexe différents, de dégager leurs variations normales et pathologiques et d'en faire, par exemple, un des éléments du diagnostic en cas de maladie ou de soupçon de maladie. Il en est de même du météorologiste et de l'agronome qui, à un endroit donné, observent les variations de la température de l'air atmosphérique aux divers mois de l'année ou de l'économiste et de l'homme d'action sociale qui observent le taux du chômage dans une profession donnée. Dans bien des cas, la connaissance des variations périodiques est pratiquement même beaucoup plus importante que celle de la moyenne de toute la période. Pour l'agronomie par exemple c'est presque toujours le cas.

§ 4. — Or, les méthodes ne sont que partiellement les mêmes dans le cas où l'on cherche l'élimination des variations périodiques et dans celui, au contraire, où l'on veut les analyser. Dans les deux cas, le problème consiste d'abord à fixer, dans chaque série empiriquement donnée, la part des variations ayant un caractère périodique déterminé (par exemple, un caractère saisonnier). Cette part une fois établie, la tâche pour ce qui concerne les fluctuations périodiques est dans le premier cas terminée ; il ne reste qu'à les éliminer, le résidu représentant l'effet des autres facteurs. Par contre, pour le second genre de recherches, ce n'est qu'une première étape. Ici, comme pour toute étude des phénomènes réels, la connaissance s'acquiert par voie de comparaison. Un second problème se pose donc ici : comment comparer les différentes séries de variations périodiques entre elles afin de dégager aussi nettement que possible leurs propriétés respectives. Ce problème, laissé généralement dans l'ombre par les recherches qui avaient pour but surtout l'élimination des fluctuations périodiques, fait précisément l'objet de la présente étude.

§ 5. — Nous commencerons par quelques définitions.

Nous appelerons *cycle* une alternance de hausse et de baisse à intervalles réguliers. Nous disons intervalles *réguliers* et pas nécessairement *égaux*, ceux-ci n'étant qu'une espèce particulière de ceux-là. L'alternance des variations peut donc avoir une durée *constante* (intervalles égaux) et une durée *variable selon une loi déterminée*. Toute-

fois, dans la présente étude, pour des raisons d'ordre pratique, nous aurons en vue uniquement les alternances constantes, auxquelles on réserve plus proprement le terme *périodicités*. Ce sont en effet les alternances les plus simples et, je crois, les seules qui dans l'état actuel de nos connaissances puissent être nettement établies.

Les variations (ou séries) accusant une périodicité, nous les appellerons *périodiques*. L'intervalle qui sépare deux hausses ou deux baisses successives marque la *période* des variations.

Le *qualificatif de temps* (annuel, mensuel, journalier, etc.) appliqué au mot « périodicité » caractérisera pour nous la durée de toute la période ; appliqué au mot « variations » (ou fluctuations), il marquera l'intervalle séparant les termes de la période les uns des autres. Ainsi, par l'expression de *périodicité journalière* (ou diurne) de la température nous désignerons l'ensemble de ses variations au cours de la journée, c'est-à-dire la période formée par des variations horaires ; par contre, par *variations journalières* de la température nous entendrons ses variations d'un jour à l'autre. De même, pour la périodicité mensuelle des cours des valeurs en bourse et pour leurs variations mensuelles (formant une périodicité annuelle), et ainsi de suite. Les variations horaires, par exemple, peuvent accuser une périodicité journalière (diurne), les variations mensuelles peuveut former une périodicité annuelle, et ainsi de suite.

Nous appellerons « séries périodiques *concrètes* » les séries des variations telles qu'elles nous sont directement données par l'observation, se trouvant encore influencées par des facteurs autres que la périodicité envisagée et comprenant chacune un certain nombre de périodes ; nous appellerons « séries périodiques *typiques* » (ou séries périodiques tout court) les séries où l'action des facteurs autres que la périodicité envisagée est éliminée et qui comprennent chacune une seule période montrant ce qu'il y a de proprement périodique (de l'espèce envisagée) dans les variations d'une série concrète (*). Dans la suite, les séries périodiques typiques seront mises entre parenthèses. Il s'ensuit de ce que nous avons dit au § précédent que dans la présente étude nous supposerons les séries typiques déjà dégagées des séries concrètes et que nous ne nous occuperons que des méthodes

---

(*) Considérant comme vraiment périodiques les séries typiques seules, on pourrait donner de la périodicité constante une définition plus rigoureuse que celle indiquée plus haut : une variation accuse une périodicité constante *si à des intervalles égaux le phénomène variable s'exprime par les mêmes nombres.*

de comparaison des séries typiques entre elles (comprenant chacune une seule période).

Selon le nombre des sommets (ou maxima) qu'accusent pendant la période les courbes qui représentent les séries périodiques typiques (en prenant chaque fois comme point de départ de la série son terme minimum), nous distinguerons des séries *monocéphales* (ayant un seul sommet) et *polycéphales* (qui en ont plusieurs). Ces dernières peuveut être subdivisées à leur tour en bicéphales, tricéphales, etc. (Pour les séries périodiques et, en général, pour les séries typiques, nous préférons ces expressions à celles de « unimodale » et « plurimodale » employées dans le cas de séries dont la dispersion est due plus ou moins au hasard). Des courbes ayant le même nombre de sommets pendant une période seront appelées *homocéphales*.

Nous appellerons *homochrones* les séries de variations dont la période a la même durée ; les séries différant entre elles par la durée de leur période seront appellées *hétérochrones*. Ainsi, les périodicités annuelles sont homochrones entre elles (de même que les périodicités journalières, etc.) ; mais une périodicité annuelle et une périodicité journalière sont entre elles hétérochrones.

Nous appellerons les séries *homonomes* si elles ont le même nombre de termes ; dans le cas contraire, nous les appellerons *hétéronomes*. Les séries peuvent être à la fois homochrones et hétéronomes (par exemple : deux séries montrant la périodicité annuelle de la température de l'air atmosphérique observée chaque mois et chaque quinzaine; donc série de 12 termes dans un cas et de 24 termes dans l'autre); elles peuvent de même être hétérochrones et homonomes (par exemple : deux séries dont l'une montre la périodicité diurne de la température observée toutes les deux heures et dont l'autre indique la périodicité annuelle de la température observée une fois par mois ; dans les deux cas on aurait une série à 12 termes).

Notons encore que la périodicité des variations n'est pas nécessairement de nature *chronologique ;* elle peut être aussi *spaciale*. On pourrait par exemple relever une périodicité dans l'ornementation d'un mur, dans le niveau d'un terrain ondulé, dans la longueur des rayons menés, à des distances angulaires determinées, du centre de la terre à divers points d'un double méridien, etc. — Dans la réalité cependant, il s'agit presque toujours de périodicités chronologiques, c'est-à-dire de variations dont la période est formée par une unité de *temps*. Notre raisonnement portera lui aussi sur des périodicités chronologiques. Toutefois, les considérations que nous développons et les conclusions

auxquelles nous aboutissons pourront, par parfaite analogie, être appliquées (en quelque sorte transposées) aussi aux périodicités spaciales (c'est-à-dire aux variations dont la période est constituée par une unité *d'espace*).

§ 6. — Des comparaisons peuvent être faites soit entre périodicités homochrones, soit aussi entre périodicités hétérochrones.

a) Beaucoup plus nombreuses et plus variées paraissent être les comparaisons possibles entre périodicités *homochrones* ; on peut, en effet, établir ici les comparaisons suivantes :

1° Comparaisons entre les variations périodiques d'un phénomène (ou caractère) donné observé dans des objets différents. Ainsi par exemple on peut comparer les fluctuations mensuelles des températures de l'eau, du sol et de l'air atmosphérique, les fluctuations mensuelles du chômage dans l'industrie du bâtiment et dans celle du vêtement, les fluctuations journalières de la fréquence des suicides parmi les hommes et parmi les femmes, etc.

2° Comparaisons entre les variations périodiques d'un phénomène donné observé à des endroits différents.

3° Comparaisons entre les variations périodiques d'un certain phénomène à des époques différentes ; ceci dans le but d'établir si les variations périodiques ont changé avec le temps. On peut par exemple comparer les fluctuations mensuelles du chômage avant et après la guerre mondiale, avant et après l'introduction de tel ou tel système d'assurance ou d'assistance aux chômeurs, de tel ou tel procédé de production, etc. ; on peut de même comparer les fluctuations horaires ou mensuelles de la température d'un certain endroit avant et après son déboisement ou son reboisement, ou encore les fluctuations horaires de la température d'un malade avant et après l'application d'un certain régime hygiénique, d'un certain médicament, et ainsi de suite. Dans tous ces exemples, il s'agit de contribuer à l'étude des effets d'un certain phénomène déterminé (guerre, système d'assurance, mode de production, déboisement, médicament, etc.). Mais ces comparaisons peuvent se faire aussi sans un pareil objectif. En effet, en comparant les séries périodiques typiques pour des époques différentes, nous vérifions dans quelle mesure notre postulat (imposé surtout par la commodité du calcul) du caractère constant d'une périodicité donnée, pouvant s'exprimer par une série typique, correspond à la réalité pour tel on tel phénomène déterminé ; nous arrivons ainsi à établir également quel correctif il y a lieu éventuellement d'appliquer à telle série considérée comme typique pour tenir compte de l'évolution

de la périodicité du phénomène avec le temps. On pourrait, dans ce but, étudier par exemple les fluctuations mensuelles de la nuptialité, de la natalité ou de la mortalité en France ou en Italie ou dans un autre pays, pour les années 1876-1880, 1901-1905 et 1926-1930. — Ces comparaisons peuvent encore avoir pour but d'établir si, et dans quelle mesure, une périodicité donnée d'un certain phénomène est fonction d'une autre périodicité (hétérochrone, à savoir de durée plus longue) du même phénomène. On peut ainsi comparer les fluctuations horaires de la température de l'air atmosphérique observées en janvier avec les mêmes fluctuations observées par exemple en juillet ou en septembre ; les fluctuations du nombre des suicides selon les jours de la semaine en été et en hiver ; les fluctuations mensuelles du chômage, du montant des affaires, des dépôts dans les banques ou dans les caisses d'épargne, etc., en années d'essor et de stagnation économiques (c'est-à-dire à divers moments du cycle économique), et ainsi de suite.

4⁰ Comparaisons entre les variations périodiques de phénomènes différents. Ainsi par exemple on peut comparer les variations mensuelles du nombre des suicides avec celles de la température de l'air atmosphérique, les variations journalières du nombre des accidents avec celles des crimes, les variations horaires des accidents de route avec celles de la circulation des véhicules, etc. Ces comparaisons peuvent avoir pour but d'établir quels sont les phénomènes qui sont plus soumis à des fluctuations périodiques d'une espèce déterminée et quels autres le sont moins et, en général, en quoi les variations périodiques d'un certain phénomène diffèrent de celles d'un certain autre. Dans certains cas, elles peuvent contribuer aussi à montrer dans quelles mesure les variations de différents phénomènes sont liées les unes aux autres, soit directement, soit indirectement.

b) Mais les comparaisons entre périodicités hétérogènes ne sont pas non plus toujours dépourvues d'intérêt. Elles peuvent, en particulier, avoir pour but d'établir laquelle des différentes périodicités auxquelles un phénomène est soumis se montre plus puissante et, éventuellement, dans quelle mesure. Pour l'agronome, par exemple, le problème peut se poser si, dans un endroit donné, les fluctuations horaires de la température de l'air sont plus fortes ou plus faibles que les fluctuations mensuelles. Pour la sociologie ainsi que pour l'action sociale le problème peut se poser si les fluctuations mensuelles du chômage dans une industrie donnée sont plus fortes ou plus faibles que ses fluctuations annuelles, selon l'état du cycle économique ; on pourrait même

établir, sous ce rapport, un classement des industries, les unes ayant un chômage soumis surtout aux fluctuations saisonnières, les autres ayant, au contraire, un chômage variant surtout avec le cycle économique. — Nos méthodes de comparaison doivent donc être telles qu'elles permettent de comparer, autant que possible, aussi des séries périodiques différant entre elles par la durée de leur période ou par le nombre de leurs termes, c'est-à-dire hétérochrones et hétéronomes.

CHAPITRE II.

Caractères et classement des séries périodiques : caractères formels ; la disposition des termes ; l'intensité des variations ; l'écartement successif.

§ 7. — Nous avons déjà remarqué que les séries périodiques peuvent différer par le nombre des termes et par la durée de leur période. Elles peuvent différer aussi sous bien d'autres rapports.

Arrêtons-nous un instant sur des différences purement *formelles*, c'est-à-dire concernant uniquement la forme numérique sous laquelle les variations sont exprimées. Sous ce rapport, les séries de variations périodiques (les concrètes comme les typiques) sont souvent classées en séries de variations *absolues* et *relatives*. On les appelle absolues lorsque l'unité de mesure dans laquelle les variations sont exprimées est indépendante de l'une quelconque de ces variations ou, en d'autres termes, si l'unité de mesure est prise en dehors de ces variations. Le plus souvent l'unité de mesure dans laquelle les variations périodiques absolues sont exprimées est la même que celle employée pour l'observation du phénomène en question (par exemple: degrés de température, montant des affaires en monnaie nationale, nombre absolu des mariages, des chômeurs, des suicides, etc.). Il arrive pourtant fréquemment que l'unité de mesure est un nombre relatif (*), mais également indépen-

---

(*) La distinction courante entre nombres absolus et relatifs est d'ailleurs elle-même toute relative, pour ne pas dire davantage ; car tout nombre est une relation entre deux grandeurs (dont l'une sert d'unité de mesure à l'autre) ; dire « nombre absolu » c'est donc comme si l'on disait « relation absolue ». Tel est cependant le pouvoir « absolu » du langage qu'on est forcé d'employer de pareilles expressions.

dant des variations périodiques du phénomène (nous avons, par exemple, pour les divers mois de l'année, le nombre des chômeurs p. 100 ouvriers assurés, le nombre des condamnations p. 100 accusés, etc.) ; dans un cas pareil, au sens de notre définition, la série n'en demeure pas moins une série absolue de variations périodiques. Il en est autrement si l'on prend comme unité de mesure des variations périodiques l'un ou l'ensemble (la moyenne) des termes de la série absolue, c'est-à-dire l'une ou l'ensemble de ces mêmes variations exprimées d'abord dans une autre unité : dans ce cas, nous avons une série de chiffres relatifs ou *d'indices* des variations périodiques, montrant l'importance de ces variations relativement à un état pris comme terme de comparaison. Le plus souvent on prend ici comme unité la moyenne arithmétique de tous les termes de la série périodique absolue ; on obtient ainsi une série de chiffres où les diverses variations sout exprimées en fractions décimales (ou en % %) de la moyenne de la période. Mais on prend quelquefois aussi comme unité de mesure un certain terme de la série absolue, par exemple : le terme maximum ou le terme minimum ou encore le terme de la série absolue qui précède immédiatement le terme considéré (*chain system*) ; dans ce dernier cas, nous obtenons une série de chiffres montrant les variations du phénomène à chaque moment (heure, jour, mois, etc.) de la période par rapport à sa valeur au moment immédiatement précédent.

Les séries absolues comme les séries relatives peuvent nous donner simplement la valeur (« absolue » ou « relative ») du phénomène considéré aux divers moments de la période (v_1, v_2, v_3, . . . v_n), mais elles peuvent montrer encore, pour ces divers moments, l'écart, la différence, entre la valeur du moment et une certaine autre ($d_1 = v_1 - v_0$; $d_2 = v_2 - v_0$; $d_3 = v_3 - v_0$; . . . $d_n = v_n - v_0$). Les premières pourraient être appelées séries des valeurs périodiques *globales*; les secondes, séries des *écarts* périodiques. Dans les séries des écarts, la valeur servant de terme de comparaison pour etablir les divers écarts peut être soit constante, soit variable. Dans le premier cas (base constante) c'est généralement la valeur moyenne (arithmetique) de toute la période qui sert de terme de comparaison, la série des écarts montrant ainsi la différence entre la valeur de chaque moment de la période et cette moyenne (différence que nous appelerons écartement *moyennal* des variations périodiques). Mais ici encore, quelque autre valeur, par exemple la valeur maxima ou la valeur minima de la période, peut également servir de terme de comparaison ; la série des écarts montre alors de combien le phénomène considéré s'écarte, aux divers moments, de son

maximum ou de son minimum. — Quant au second cas, où le terme de comparaison est variable, c'est presque toujours la valeur du moment précédant immédiatement le moment envisagé qui est prise comme terme de comparaison (soit : $e_1 = v_1 - v_n$; $e_2 = v_2 - v_1$; $e_3 = v_3 - v_2$; $e_n = v_n - v_{n-1}$). Les séries des écarts périodiques (absolues comme relatives) peuvent ainsi à leur tour être classées en séries à base de comparaison fixe et en séries à base mobile. Si (comme c'est presque toujours le cas) ces dernières marquent les écarts entre deux moments immédiatement successifs, on a des séries périodiques d'écarts que nous dirons *successifs*, séries construites d'après le système dit de chaîne (*chain system*).

Notons encore que les séries *relatives* des écarts périodiques peuvent indiquer l'importance relative de chaque écart non seulement en fractions (ou % %) de la valeur ayant servi de base pour l'établissement des écarts absolus (c'est-à-dire non seulement en % de la valeur moyenne, maxima ou minima de la période, en cas de série à base fixe et en % du terme immédiatement précédent, en cas de séries du système de chaîne), mais aussi en fractions de l'écart moyen de toute la série absolue correspondante, de l'écart arithmetique moyen, ou de l'écart quadratique moyen (*standard deviation*) en cas de série à base fixe ou de la différence moyenne entre les termes successifs (« *indice d'oscillation* », *de Gini, soit la moyenne des écarts successifs*) en cas de série du système de chaîne.

D'après les principales formes numériques sous lesquelles les variations sont représentées, les séries périodiques peuvent ainsi être classées de la façon suivante :

Principales formes de séries périodiques :

Séries :	Séries de valeurs globales	Séries d'écarts			
		à base fixe			à base mobile :
	(v)	écarts de la valeur moyenne (d)	écarts de la valeur maxima (d')	écarts de la valeur minima (d'')	système de chaîne (e)
absolues.	v_1	$d_1 = v_1 - v_m$	$d'_1 = v_1 - v_{\max}$	$d''_1 = v_1 - v_{\min}$	$e_1 = v_1 - v_n$
	v_2	$d_2 = v_2 - v_m$	$d'_2 = v_2 - v_{\max}$	$d''_2 = v_2 - v_{\min}$	$e_2 = v_2 - v_1$
	v_3	$d_3 = v_3 - v_m$	$d'_3 = v_3 - v_{\max}$	$d''_3 = v_3 - v_{\min}$	$e_3 = v_3 - v_2$
	⋮	⋮	⋮	⋮	⋮
	v_n	$d_n = v_n - v_m$	$d'_n = v_n - v_{\max}$	$d''_n = v_n - v_{\min}$	$e_n = v_n - v_{n-1}$

	par rapport à la valeur				par rapport à la valeur				par rapport à la valeur	
	moyenne	maxima	minima	du moment précédent (système de chaîne)	moyenne	de l'écart quadratique moyen	maxima	minima	du moment précédent	de l'écart quadratique moyen
relatives.	$v_1 : v_m$	$v_1 : v_{\max}$	$v_1 : v_{\min}$	$v_1 : v_n$	$d_1 : v_m$	$d_1 : \sigma$	$d'_1 : v_{\max}$	$d''_1 : v_{\min}$	$e_1 : v_n$	$e_1 : \varsigma$
	$v_2 : v_m$	$v_2 : v_{\max}$	$v_2 : v_{\min}$	$v_2 : v_1$	$d_2 : v_m$	$d_2 : \sigma$	$d'_2 : v_{\max}$	$d''_2 : v_{\min}$	$e_2 : v_1$	$e_2 : \varsigma$
	$v_3 : v_m$	$v_3 : v_{\max}$	$v_3 : v_{\min}$	$v_3 : v_2$	$d_3 : v_m$	$d_3 : \sigma$	$d'_3 : v_{\max}$	$d''_3 : v_{\min}$	$e_3 : v_2$	$e_3 : \varsigma$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$v_n : v_m$	$v_n : v_{\max}$	$v_n : v_{\min}$	$v_n : v_{n-1}$	$d_n : v_m$	$d_n : \sigma$	$d'_n : v_{\max}$	$d''_n : v_{\min}$	$e_n : v_{n-1}$	$e_n : \varsigma$

n = nombre des termes ; v_m = moyenne arithmétique des valeurs globales ; σ = écart moyennal quadratique moyen $\left(= \sqrt{\frac{\Sigma d^2}{n}}\right)$; ς = écart successif quadratique moyen $\left(= \sqrt{\frac{\Sigma e^2}{n}}\right)$.

Parmi les séries absolues, ce sont les séries des écarts qui expriment ce qu'il y a de proprement périodique dans les variations d'un phénomène considéré, les séries des valeurs globales contenant outre les variations périodiques, encore un élément constant (généralement la moyenne de toute la période), indépendant de la situation dans la période. Cependant, ce sont les séries des valeurs globales qui constituent le point de départ d'où les séries des écarts sont tirées ; d'une série de valeurs globales ou peut ainsi toujours aboutir à toutes les séries des écarts tandis que d'après une série d'écarts seuls on ne peut pas reconstruire la série des valeurs globales. Ce sont aussi les séries des valeurs globales que nous prendrons comme point de départ de notre raisonnement, tout en ne perdant jamais de vue que seuls les écarts sont l'expression de la périodicité des phénomènes.

§ 8. — Mais les différences entre séries périodiques envisagées dans le § précédent sout toutes d'ordre purement formel ; de plus, elles dépendent de nous, de la forme dont nous revêtons les séries. Si nous voulons comparer des séries exprimées de façons différentes, nous pouvons, quand c'est nécessaire, les ramener à une forme unique. Ces différences n'empêchent pas de comparer les séries et, d'un autre côté, si elles étaient les seules différences entre certaines séries, ces séries seraient au fond identiques et leur comparaison deviendrait sans objet. Inutile de dire que dans la réalité les séries périodiques diffèrent toujours plus ou moins l'une de l'autre par la nature même de leurs variations. En comparant même des séries à périodicité constante, homochrones et homonomes, on constate qu'elles peuvent encore différer entre elles, surtout :

a) par la disposition des termes,
b) par l'intensité des variations,
c) par leur degré de concentration,
d) par leur degré de précipitation et,
e) par leur degré de symétrie.

Reprenons chacun de ces points separément. Commençons par les différences entre séries périodiques provenant d'une disposition différente de leurs termes.

§ 9. — Nous avons déjà marqué en passant (§ 5) que les séries peuvent être monocéphales ou polycéphales, c'est-à-dirc qu'elles peuvent différer par le *nombre de sommets* qu'elles présentent pendant une

période. Lorsqu'une série (ou, si l'on préfère, une courbe) accuse plusieurs sommets, le maximum le plus élevé et le minimum le plus bas seront appelés par nous maximum et minimum *principaux* (*) par opposition aux autres qui seront appelés (maxima et minima) *secondaires*. Mais ces derniers ne sont des maxima (et des minima) que par rapport aux termes immédiatement voisins. Si la place des termes change, un maximum (ou minimum) secondaire peut disparaître, un nouvau peut surgir, même si les termes eux-mêmes ne changent pas. C'ést donc uniquement de la disposition des termes dans la série que dépend le nombre des sommets de la série, en particulier son caractère monocéphalique ou polycéphalique (le nombre des sommets ne peut cependant pas dépasser $\frac{n}{2}$ si le nombre des termes (n) est pair et $\frac{n - 1}{2}$ s'il est impair). Ainsi par exemple, les mêmes sept termes peuvent donner, selon leur disposition (**) :

des séries monocéphales comme $(\overline{18}, 15, 12, 10, 7, 6, \underline{2})$,

des séries bicéphales comme $(\overline{18}, 7, \underline{6}, 10\ \overline{15}, \underline{2}, 12)$ ou

des séries tricéphales comme $(\overline{18}, \underline{12}, \overline{15}, 10, \underline{2}, \overline{7}, 6)$. (***)

(*) On les appelle généralement *absolus*. A notre avis, cependant, on abuse vraiment trop du terme « absolu ». Nous aurons encore l'expression « valeur absolue » dans le sens de l'indépendance du signe (+ ou —). Voici donc la quatrième fois que nous aurions au cours de cette étude le mot « absolu » (deux fois au § 7), et chaque fois dans un autre sens. Si le mot « absolu » devait être pris comme synonyme d'indépendant d'*une* autre chose envisagée à un moment donné, ce terme pourrait en effet revêtir autant de sens qu'il y a de choses à envisager, c'est-à-dire un nombre pour ainsi dire infini de sens différents qui, se rencontrant ensuite, pourraient provoquer des confusions et des cacophonies sans nombre (par exemple : la valeur absolue du minimum absolu des variations absolues du nombre absolu, ou relatif, des chômeurs, etc.). On se demande si le mot absolu ne devrait pas être pris plutôt dans son sens propre, dans le sens d'indépendant de *toute* chose et comme tel, être complètement banni du vocabulaire scientifique (voy. aussi la note précédente, au § 7).

(**) Nous mettons un trait audessus des chiffres pour marquer les maxima et au-dessous d'eux pour signaler les minima.

(***) Il est a remarquer à ce sujet que dans les séries chronologiques la distribution des termes n'est evidemment pas arbitraire, mais fixée par leur succession dans le temps.

Or le nombre des sommets d'une série périodique doit évidemment être pris en considération quand on compare les séries entre elles. Il se peut en effet que telle industrie, par exemple, accuse, une seule morte saison (supposons en hiver) et que telle autre en accuse deux (disons au milieu de l'hiver et au milieu de l'été) ; il se peut que la nuptialité accuse, au cours de l'année un seul maximum dans un certain pays, deux maxima dans un autre et trois dans un troisième ; il se peut que la mortalité des différents groupes d'âge accusent un nombre différent de maxima au cours de l'année, et ainsi de suite. Pour l'étiologie des variations d'un phénomène considéré ainsi que pour l'action pratique, il n'est évidemment pas indifférent de savoir quel est le nombre de hausses et de baisses qu'il subit au cours d'une période donnée.

Mais avant tout la disposition des termes dans la série détermine la *situation* des maxima et des minima, le *moment* où le phénomène est à son maximum et à son minimum. La comparaison des séries périodiques sous ce rapport peut être aussi d'un grand intérêt. Et ce n'est pas seulement la coïncidence des moments d'intensité maxima (et minima) pour des séries différentes qui peut être instructive et utile ; leur non-coïncidence ne l'est pas moins. Ainsi, par exemple, l'étude des intervalles qui séparent les maxima (et les minima) de température annuelle de l'air atmosphérique et du sol observés à différentes hauteurs et profondeurs peut être extrêmement instructive pour la géophysique, pour la botanique ainsi que pour les applications pratiques sur le terrain agricole ; la situation des maxima et des minima de chômage saisonnier dans différentes branches économiques ou professions, particulièrement dans le cas de non-coïncidence de ces « mortes saisons » peut être très importante pour l'organisation de la lutte contre le chômage, et ainsi de suite.

La disposition des termes dans la série *contribue* aussi à déterminer certains autres caractères de la série, comme nous le verrons dans la suite.

§ 10. — L'*intensité* des variations périodiques trouve son expression surtout :

1⁰) dans l'ensemble des écarts entre les divers termes de la série et leur moyenne générale, soit dans *l'écartement moyennal de tous les termes* de la série ;

2⁰) dans la différence entre le maximum et le minimum de la série, c'est-à-dire dans *l'amplitude des variations de la série ;*

3°) dans les différences entre les divers termes, soit entre tous les termes de la série abstraction faite de leur ordre (*différence moyenne*), soit entre les termes successifs (*écartement successif*).

Inutile d'insister sur l'importance qu'a, pour l'étude des variations périodiques, l'ensemble des écarts entre les termes de la série et leur moyenne générale. Inutile aussi d'insister sur les procédés de mensuration de cet écartement moyennal des termes de la série. Comme d'habitude, on peut notamment envisager soit l'écart quadratique moyen $\left(\sigma = \sqrt{\frac{\Sigma d^2}{n}}\right)$, soit simplement l'écart arithmétique moyen $\left(\text{que nous désignerons désormais par } d_m = \frac{\Sigma d}{n}\right)$. — Pour des raisons analogues, nous ne nous arrêterons pas sur la « différence moyenne » (ou « indice d'oscillation »).

Quant à l'amplitude des variations, elle est donnée par la différence entre le terme maximum et le terme minimum de la série, c'est-à-dire entre le terme accusant le plus grand écart en plus de la moyenne de la série et le terme accusant le plus grand écart en moins de la moyenne. Si nous désignons l'amplitude des variations par α, les termes de la série par v, les écarts moyennaux positifs par x et les valeurs absolues des écarts moyennaux négatifs par z, nous pouvons ainsi écrire :

$$\alpha = v_{\max} - v_{\min} = (v_{\text{moy}} + x_{\max}) - (v_{\text{moy}} - z_{\max}) = x_{\max} + z_{\max}. \quad [1]$$

Le maximum ($v_{\max}$) et le minimum ($v_{\min}$) de la série seront appelés par nous les termes *extrêmes* de la série périodique ; tous les autres seront désignés sous le nom de termes *intermédiaires*.

Dans les observations où la dispersion se fait plus ou moins au hasard, les termes extrêmes, accusant les écarts maxima à la moyenne soit en plus soit en moins, peuvent être en un sens considérés comme dépourvus de signification ; certaines méthodes statistiques généralement admises (comme par exemple la médiane des écarts ou écart probable, les quartiles, etc.) sont même basées sur la négligence complète de la valeur des écarts dépassant certaines limites. Il en est autrement des fluctuations se répétant périodiquement : les écarts maxima de la moyenne ou, en d'autres termes, la valeur maxima et la valeur minima des variations périodiques typiques d'un phénomène donné marquent les points où la périodicité

se manifeste avec le plus d'intensité. Il s'agit là de ces *instantiae ostensivae* que BACON a mis au premier rang des faits capables d'éclairer la recherche scientifique. Ce sont en même temps aussi les points critiques qui dans bien des cas peuvent être déterminants pour l'action pratique (en agronomie, en médecine, en politique et ailleurs). Nous verrons dans la suite que non seulement pour eux-mêmes, mais encore pour l'élaboration de certaines autres méthodes de mensuration des variations périodiques, les termes extrêmes présentent un grand intérêt.

L'intensité des variations trouve son expression aussi dans les différences entre les termes de la série. Seulement, cet aspect des variations périodiques est-il indépendant des caractères que nous avons déjà marqués ? Pour la différence moyenne entre tous les termes, ses relations avec l'écartement moyen arithmétique ou quadratique ont été déjà étudiées (cfr. surtout C. GINI : *Variabilità e Mutabilità*, 1912 — G. PIETRA : *Le relazioni tra gli indici di variabilità.* Atti del R. Istituto Veneto di Scienze Lettere ed Arti, 1914). La relation demeure constante pour les courbes d'un type determiné et change naturellement pour des courbes de type différent. Nous fixerons notre attention spécialement sur l'écartement successif. Ne se trouve-t-il pas entièrement ou partiellement déjà déterminé par la disposition des termes, par l'écartement moyennal et l'amplitude des variations ? Pour y voir clair, nous allons examiner de plus près l'écartement successif des termes dans des séries périodiques.

§ 11. — Nous examinerons l'écartement successif des termes des séries périodiques séparément dans le cas de séries monocéphales et dans le cas de séries polycéphales.

A. — *Séries périodiques monocéphales.*

PROPOSITION. — *Dans toute série périodique monocéphale, la somme des valeurs absolues des écarts successifs est égale au double de l'amplitude des variations de la série.*

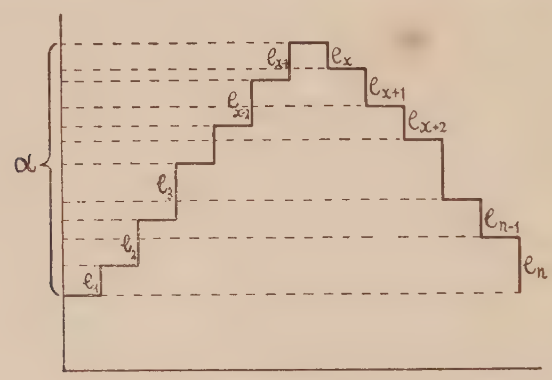
En d'autres termes, désignant les valeurs absolues des écarts entre les termes successifs (y compris naturellement l'écart entre le premier et le dernier terme) de la série périodique (typique) par e et l'amplitude des variations par α, nous pouvons dire :

$$\Sigma e = 2\alpha \qquad [2]$$

Prenons en effet une série périodique monocéphale, désignons le minimum de la série par v_1, le maximum par v_k et les termes intermédiaires par v_2, $v_3 \ldots v_{k-1}, v_{k+1}, \ldots v_{n-1}, v_n$; la série sera donc ascendante de v_1 à v_k et descendante de v_k à v_1 (à travers v_n). La somme des valeurs absolues des écarts successifs entre les termes de cette série sera égale à :

$$\Sigma e = [(v_2 - v_1) + (v_3 - v_2) + (v_4 - v_3) + \ldots + (v_{k-1} - v_{k-2}) +$$
$$+ (v_k - v_{k-1})] + [(v_k - v_{k+1}) + (v_{k+1} - v_{k+2}) + (v_{k+2} - v_{k+3}) + \ldots +$$
$$+ (v_{n-2} - v_{n-1}) + (v_{n-1} - v_n) + (v_n - v_1)] = [- v_1 + v_k] + [v_k - v_1] =$$
$$= \alpha + \alpha = 2\alpha\,.$$

On peut d'ailleurs d'emblée se rendre compte de l'exactitude de cette proposition par le simple graphique ci-dessous représentant sché-



Graphique I.

matiquement une courbe périodique (typique) monocéphale. Il est évident que la somme des valeurs absolues des écarts successifs $e_1 + e_2 + e_3 + \ldots + e_{k-2} + e_{k-1} = \alpha$ de même que la somme $e_k + e_{k+1} + e_{k+2} + \ldots + e_{n-1} + e_n = \alpha$; la somme des deux est donc égale à 2α.

§ 12. — Corollaires.

1°) *Pour autant qu'une série périodique demeure monocéphale, la somme des valeurs absolues des écarts des termes successifs ne dépend ni de la place des termes, ni des termes intermédiaires eux-mêmes, ni*

du nombre des termes, mais uniquement de la différence des termes extrêmes.

2°) *L'écart successif arithmétique moyen de séries périodiques monocéphales ayant une amplitude donnée* (α) *est indépendant de la place des termes ainsi que de la valeur des termes intermédiaires et est inversement proportionnel au nombre des termes* (n).

En effet, désignant l'écart successif arithmétique moyen par e_m, nous avons :

$$e_m = \frac{\Sigma e}{n} = \frac{2}{n}\alpha \,. \qquad [3]$$

En d'autres termes, *pour des séries périodiques monocéphales homonomes ayant la même amplitude, l'écart successif arithmétique moyen est constant.*

Voici quelques exemples concrets qui illustrent ces déductions :

Séries	Σe	e_m
$(\overline{18}, 15, 12, 10, 7, 6, 5, \underline{2})$	32	4

Changeons la place des termes :

$(\overline{18}, 12, 7, 5, \underline{2}, 6, 10, 15)$	32	4

Changeons les termes intermédiaires :

$(\overline{18}, 5, 4, 3, 3, \underline{2}, 16, 17)$	32	4

Changeons le nombre des termes :

$(\overline{18}, 4, \underline{2}, 9)$	32	8
$(\overline{20}, 17, 12, 11, 6, \underline{4}, 5, 7, 8, 11, 13, 15)$	32	2,67

B. — *Séries périodiques polycéphales.*

§ 13. — Définition : Nous appellerons *amplitude partielle* (α') la différence entre un maximum quelconque ($v_{max'}$) et le minimum suivant ($v_{min'}$), soit :

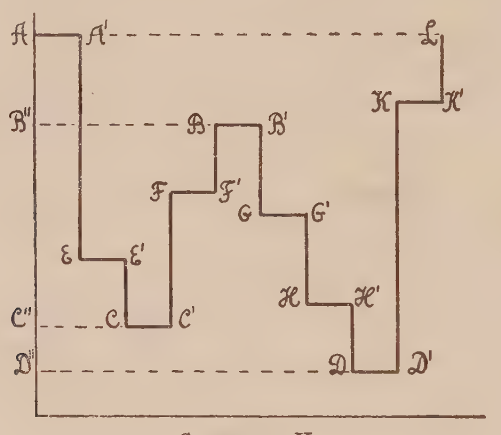
$\alpha'_1 = v_{max'(1)} - v_{min'(1)}$; $\alpha'_2 = v_{max'(2)} - v_{min'(2)}$, et ainsi de suite.

Si le minimum qui suit le maximum principal de la série est lui aussi le minimum principal de la série, cette amplitude partielle

coïncide évidemment avec l'amplitude générale de la série (ou avec son amplitude tout court).

Proposition. — *Dans toute série périodique polycéphale, la somme des valeurs absolues des écarts successifs est égale au double de la somme des amplitudes partielles de la série.*

On peut s'en rendre compte en divisant la série en autant de parties qu'elle compte de sommets et en raisonnant d'une façon analogue à celle appliquée aux séries monocépales (§ 11). On peut s'en rendre compte également en raisonnant d'après un graphique schématique (d'une courbe bicéphale) comme celui-ci :



Graphique II.

La somme des valeurs absolues des écarts successifs de la courbe périodique (typique) bicéphale $A\ C\ B\ D\ L$ est égale à :

$$\Sigma e = (A'\,E + E'\,C) + (C'\,F + F'\,B) + (B'\,G + G'\,H + H'\,D) +$$

$$+ (D'\,K + K'\,L) = A\,C'' + B''\,C'' + B''\,D'' + A\,D''\,.$$

Notons que $A\,C''$ représente la première amplitude partielle de notre courbe périodique que nous pouvons désigner par α'_1 ; $B''\,D''$ représente la seconde amplitude partielle de cette courbe, donc α'_2 ;

$$B''\,C'' = A\,C'' - A\,B'' = \alpha'_1 - A\,B'' \ ; \ A\,D'' = B''\,D'' +$$

$$+ A\,B'' = \alpha'_2 + A\,B''\,.$$

Nous pouvons donc écrire :

$$\Sigma\, e = \alpha'_1 + (\alpha'_1 - A\,B'') + (\alpha'_2 + A\,B'') + \alpha'_2 = 2\,\alpha'_1 + \\ + 2\,\alpha'_2 = 2\,(\alpha'_1 + \alpha'_2)\,,$$

soit d'une façon générale :

$$\Sigma\, e = 2\,\Sigma\,\alpha'\,. \qquad [4]$$

Ajoutons encore que, puisque $\alpha' = v_{\max'} - v_{\min'}$, il est évident que $\Sigma\,\alpha' = \Sigma\,(v_{\max'} - v_{\min'}) = \Sigma\, v_{\max'} - \Sigma\, v_{\min'}$;
nous pouvons donc écrire aussi :

$$\Sigma\, e = 2\,(\Sigma\, v_{\max'} - \Sigma\, v_{\min'}) \qquad [5]$$

c'est-à-dire que *la somme des valeurs absolues des écarts des termes successifs est dans toute série périodique égale au double de la différence entre la somme des maxima et la somme des minima.*

§ 14. — Corollaires :

1°) Comme les maxima et les minima secondaires peuvent changer avec le changement de place des termes de la série (§ 9), *la somme des écarts successifs d'une série polycéphale dépend de la disposition des termes dans la série.*

2°) *Pour autant que les maxima et les minima des séries périodiques demeurent les mêmes, la somme des écarts successifs est constante ;* elle ne dépend alors ni de la place des termes, ni des termes intermédiares eux-mêmes, ni de leur nombre.

3°) *L'écart successif arithmétique moyen* est égal à :

$$e_m = \frac{2\,\Sigma\,\alpha'}{n} = \frac{2}{n}\,(\Sigma\, v_{\max'} - \Sigma\, v_{\min'})\,. \qquad [6]$$

Pour une somme d'amplitudes partielles donnée $(\Sigma\,\alpha')$, l'écart successif arithmétique moyen est ainsi inversement proportionnel au nombre des termes et ne dépend pas des autres éléments de la série.

4°) Pour une amplitude générale (α) donnée, la série *monocéphale* a la somme (et la moyenne) des valeurs absolues des écarts suc-

cessifs *minima*. (La série monocéphale peut être considérée comme un cas particulier de séries polycéphales où les amplitudes partielles autres que la différence entre le maximum et le minimum principaux sont égales à zéro).

Voici, à titre d'illustration, quelques exemples concrets :

	Σe	e_m
Série monocéphale :		
$(\overline{18}, 15, 10, 7, 5, \underline{2}, 6, 12)$. .	32	4 — (coroll. 4)
Séries bicéphales :		
$(\overline{18}, 5, \underline{2}, \overline{15}, \underline{6}, 7, 10, 12)$. .	50	6,25 (coroll. 2)
$(\overline{18}, 10, \underline{6}, 10, \overline{15}, 12, 5, \underline{2})$.	50	6,25 (» 2)
$(\overline{18}, 13, 8, 4, \underline{2}, 11, \overline{15}, \underline{6})$. .	50	6,25 (» 2)
$(\overline{18}, 15, 6, \underline{5}, 7, \overline{10}, \underline{2}, 12)$. .	42	5,25 (» 1)
$(18, \underline{6}, \overline{15}, \underline{2})$	50	12,50 (» 2 et 3)

Et ainsi de suite.

CHAPITRE III.

Caractères et classement des séries périodiques (*suite*) : concentration, précipitation et symétrie des variations.

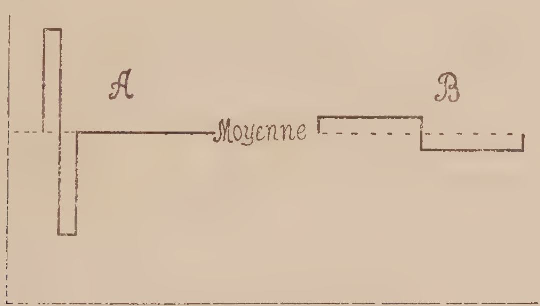
§ 15. — Deux séries accusant la même somme d'écarts moyennaux (dont l'ensemble des termes s'écarte de la moyenne de la série dans une mesure égale) peuvent cependant être très différentes l'une de l'autre selon la répartition plus ou moins égale, ou plus ou moins inégale, de cette somme des écarts sur les divers termes de la série.

Nous pouvons faire ici une observation analogue à celle faite plus haut au sujet de l'amplitude des variations. Dans les séries où les écarts peuvent être considérés comme accidentels, la distribution des écarts se rapproche de la courbe normale et les différences qui sous ce rapport existent entre diverses séries réelles peuvent

être regardées comme relativement secondaires. Il n'en est pas de même des séries de variations *typiques*, dont les écarts sont l'effet de causes systématiques. Ici la fréquence différente des divers écarts que l'on observe dans différentes séries n'est nullement négligeable, mais caractérise ces séries au même titre que la somme de leurs écarts ou leur écart moyen.

Pour être tout à fait clair, prenons un exemple schématique. Prenons deux séries à 12 termes, A et B, s'exprimant par les chiffres suivants :

		En moyenne
	Séries données :	
A	(16, 4, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10)	10
B	(11, 11, 11, 11, 11, 11, 9, 9, 9, 9, 9, 9)	10
	Écarts moyennaux :	
A	(+ 6, — 6, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	1
B	(+ 1, + 1, + 1, + 1, + 1, + 1, — 1, — 1, — 1, — 1, — 1, — 1)	1



GRAPHIQUE III.

Nous avons ainsi deux espèces de courbes (A et B) qui, en moyenne accusent des variations de dimensions égales ($d_m = 1$), la somme des valeurs absolues de leurs écarts moyennaux étant la même ($\Sigma d = 12$). Seulement, dans l'une d'elles (A), les dimensions des écarts sont beaucoup plus considérables ($= \pm 6$), mais ces écarts ne sont observés que pour peu de termes (deux), tandis que dans l'autre (B) les écarts sont beaucoup moins importants ($= \pm 1$), mais sont constatés pour beaucoup plus de termes (douze). Dans le premier cas (A), nous dirons que les variations sont plus *concentrées ;* dans l'autre (B), qu'elles sont plus *étendues, ou prolongées, ou nivelées.*

Le degré de concentration (ou de nivellement) des écarts signifie ainsi la répartition plus ou moins inégale (ou plus ou moins égale) de la somme des écarts moyennaux de la série sur ses divers termes.

Si, dans les deux séries A et B, nous comparons chaque écart moyennal (y compris les écarts 0) avec l'écart moyennal moyen de la série (qui dans les deux cas est égal à 1), nous obtenons pour chacune d'elles une série d'*écarts moyennaux du second ordre* (différences entre les écarts réels et l'écart moyen de la série, soit : $\delta = d - d_m$ en désignant par d les valeurs absolues des écarts moyennaux du premier ordre et par δ ceux du second ordre) ; voici ces séries :

		En moyenne (δ_m)
Écarts moyennaux du second ordre (δ) :		
A	(+ 5, + 5, — 1, — 1, — 1, — 1, — 1, — 1, — 1, — 1, — 1, — 1)	$\frac{20}{12} = 1{,}67$
B	(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0

Dans la série B, où la somme des écarts moyennaux (du premier ordre) est répartie d'une façon *égale* entre tous les termes de la série, où la concentration de l'écartement moyennal est *nulle* (et où son nivellement est maximum), les écarts moyennaux du second ordre sont tous égaux à O ; par contre, dans la série A, qui accuse des écarts moyennaux très inégaux, où, en d'autres termes, la fluctuation se trouve très concentrée (en tout, sur deux termes de la série), l'extension des fluctuations étant très limitée, les écarts moyennaux du second ordre sont assez importants. Pour le même écartement moyennal moyen (du premier ordre), mais à un degré différent de concentration et de nivellement (ou extension) des fluctuations, répond ainsi une différence d'écartement moyennal du second ordre.

§ 16. — Notons bien que la différence de concentration (et de nivellement) des écarts des différentes séries n'est nullement le simple résultat de la différence d'amplitude de leurs variations (comme on pourrait peut-être croire d'après notre graphique). Il se peut que deux séries aient le même écartement moyennal arithmétique moyen (soit la même somme des écarts moyennaux) et la même amplitude des variations et que la concentration de leurs variations soit quand même différente. Prenons, en effet, deux séries comme les suivantes :

		En moyenne
	Séries données :	
A_1	(16, 16, 4, 4, 10, 10, 10, 10, 10, 10, 10, 10)	10
B_1	(16, 14, 12, 9, 9, 8, 8, 4, 10, 10, 10, 10)	10
	Écarts moyennaux :	
	du 1^er^ ordre (d) :	
A_1	(+ 6, + 6, — 6, — 6, 0, 0, 0, 0, 0, 0, 0, 0)	2
B_1	(+ 6, + 4, + 2, — 1, — 1, — 2, — 2, — 6, 0, 0, 0, 0)	2
	du 2^d^ ordre (δ) :	
A_1	(+ 4, + 4, + 4, + 4, — 2, — 2, — 2, — 2, — 2, — 2, — 2, — 2)	$\frac{32}{12} = 2{,}67$
B_1	(+ 4, + 2, 0, — 1, — 1, 0, 0, + 4, — 2, — 2, — 2, — 2)	$\frac{20}{12} = 1{,}67$

Dans les deux séries, la somme des valeurs absolues des écarts (du 1^er^ ordre) est la même ($\Sigma\, d = 24$) de même que leur écart arithmétique moyen ($d_m = 2$) ; l'amplitude des variations est aussi la même ($\alpha = 12$) ; et pourtant dans la série A_1 les fluctuations sont plus concentrées (en tout sur quatre termes) que dans la série B_1 (où elles s'étendent sur huit termes) ; aussi, l'écartement du second ordre est-il plus fort pour la série A_1 que pour B_1 (2,67 contre 1,67).

Remarquons encore que dans nos observations relatives à la concentration et à l'extension (ou nivellement) des fluctuations nous n'avons pas fait intervenir de considérations basées sur le caractère périodique des séries. Ces observations s'appliquent donc aux séries statistiques en général, aux séries ordinaires comme aux séries périodiques. Ce sont pourtant les séries périodiques qui nous intéressent ici en première ligne. Or, le degré de concentration ou d'extension (nivellement) des fluctuations périodiques n'est nullement une chose indifférente dans l'étude de la périodicité des phénomènes, ni au point de vue théorique ni au point de vue pratique. Pour le météorologiste ou l'agronome, en effet, ce n'est pas seulement l'écart annuel *moyen* des températures mensuelles (par rapport à la température moyenne de l'année) qui compte, mais aussi (si ce n'est pas surtout) le degré d'égalité ou d'inégalité de la distribution de la somme des écarts sur les divers mois de l'année, c'est-à-dire la concentration ou le nivellement (extension) des écarts saisonniers ; en d'autres termes, il importe aussi de savoir si les écarts saisonniers se rapprochent d'une

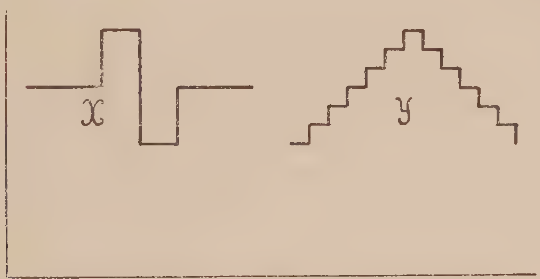
courbe du type A ou de celle du type B (du § 15). — Il peut en être de même pour le physiologiste ou le médicien en ce qui concerne les fluctuations horaires de la température du corps humain. De même l'économiste et l'homme de politique sociale ne pourront pas assimiler les cas de deux industries dont l'une (type A) subit une morte saison augmentant son chômage normal de 150 p.100 pendant un mois de l'année tandis que l'autre(type B) accuse un chômage accru de 25 p. 100 durant six mois sur douze ; la gravité que peut revêtir le phenomène à un moment donné ainsi que la durée pendant laquelle il cause des soucis sont très différents dans les deux cas et les moyens d'action ne sauraient être identiques.

§ 17. — Il importe aussi de distinguer les courbes (ou séries) des variations périodiques selon leur allure plus ou moins brusque, précipitée ou, au contraire, graduelle, ondulatoire. Mais qu'est-ce qui détermine ce caractère des séries ?

La précipitation ou la gradation d'une série dépend évidemment de l'*écartement successif* de ses termes. Cependant, un peu de réflexion suffit pour nous montrer que pour des courbes homocéphales (ayant le même nombre de sommets), ce n'est pas l'importance globale de cet écartement (qui d'ailleurs s'exprimera par des chiffres très différents selon l'unité de mesure adoptée), mais sa *répartition plus ou moins égale ou inégale* entre les divers termes qui détermine l'allure plus ou moins graduelle, ondulatoire ou au contraire précipitée de la série. Des courbes périodiques homocéphales, homonomes et ayant les mêmes amplitudes, ayant par suite aussi la même somme des valeurs absolues des écarts successifs et le même écart successif arithmétique moyen (§§ 12 et 14), peuvent en effet accuser une allure très différente au point de vue de la précipitation ou de la gradualité de leurs variations selon que cette somme des écarts successifs est répartie plus ou moins également entre les termes de chaque série (ou ce qui revient au même, selon que les écarts successifs réels diffèrent plus ou moins de l'écart successif moyen), c'est-à-dire selon l'*écartement successif du second ordre*. — Voici, à titre d'exemple, deux séries monocéphales, homonomes et ayant la même amplitude dont l'une (X) accuse des variations d'une allure précipitée tandis que l'autre (Y) a une allure très ondulatoire, graduelle :

	En moyenne
Séries données :	
X (10, 10, 10, 10, 13, 13, 7, 7, 10, 10, 10, 10)	10
Y (7, 8, 9, 10, 11, 12, 13, 12, 11, 10, 9, 8)	10
Écarts successifs	
du 1^er ordre (*e*) :	
X (0, 0, 0, 0, + 3, 0, — 6, 0, + 3, 0, 0, 0)	1
Y (— 1, + 1, + 1, + 1, + 1, + 1, + 1, — 1, — 1, — 1, — 1, — 1)	1
du 2^d ordre (*) (ε) :	
X (0, 0, 0, 0, + 3, — 3, — 6, + 6, + 3, — 3, 0, 0)	2
Y (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)	0

Dans les deux séries, l'amplitude des variations et, par suite, la somme des valeurs absolues des écarts successifs (du premier ordre) est la même ($\alpha = 13 - 7 = 6$; $\Sigma e = 12$) ; l'écart successif arithmétique moyen du premier ordre est donc aussi le même dans les deux cas ($e_m = \frac{12}{12} = 1$). Mais l'écart successif moyen du second ordre (ε_m) est, pour la série X, très différent de celui de la série Y : dans la



GRAPHIQUE IV.

série X, où les écarts entre termes voisins diffèrent beaucoup les uns des autres et qui accuse par suite des hausses et des baisses assez précipitées, l'écart successif moyen du second ordre est égal à 2 ; par contre, dans la série Y, dont les variations accusent une parfaite gradation, chaque terme différant du terme voisin de la même quan-

(*) L'écartement successif du second ordre mériterait une étude plus détaillée ; nous sommes cependant obligé d'y renoncer ici pour ne pas nous éloigner de notre sujet proprement dit.

tité (en valeur absolue), l'écart successif moyen du second ordre est égal à zéro.

Pour des séries homocéphales, la précipitation et la gradualité des variations constituent ainsi une fonction de leur écartement successif qui est parfaitement analogue à celle que la concentration et le nivellement constituent par rapport à leur écartement moyennal.

§ 18. — Les séries périodiques peuvent différer entre elles encore selon le degré de leur *symétrie ou asymétrie.*

Bien que le problème de la symétrie ait été déjà à maintes reprises examiné par des statisticiens, il nous semble pourtant loin d'être suffisamment élucidé. Nous ne croyons donc pas inutile de nous y arrêter un peu plus longuement. Commençons par la *définition* même de la symétrie.

Une série est symétrique si tous les termes équidistants de son terme central (*) *marquent par rapport à celui-ci des différences soit égales, soit ayant des valeurs absolues égales et des signes toujours opposés.* — Ainsi, une série (où le terme central est désigné par a_0 et dont le nombre des termes est égal à $2n + 1$)

$$a'_n\,,\; a'_{n-1}\,,\; \ldots\; a'_2\,,\; a'_1\,,\; a'_0\,,\; a''_1\,,\; a''_2\,,\; \ldots\; a''_{n-1}\,,\; a''_n$$

est symétrique si :

$$a'_1 - a_0 = a''_1 - a_0\,,\; a'_2 - a_0 = a''_2 - a_0\,,\; \ldots\; a'_n - a_0 = a''_n - a_0\,,$$

ou encore si :

$$(a'_1 - a_0) = -(a''_1 - a_0)\,,\; (a'_2 - a_0) = -(a''_2 - a_0)\,,\; \ldots$$
$$(a'_n - a_0) = -(a''_n - a_0)\,.$$

En désignant ces différences entre les termes de la série et le terme central par η, nous pouvons donc dire autrement qu'une série est symétrique si elle peut être exprimée sous la forme suivante :

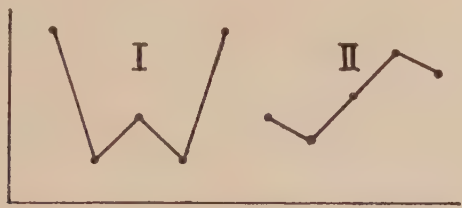
$$(a_0 + \eta_n)\,,\; (a_0 + \eta_{n-1})\,,\; \ldots\,,\; (a_0 + \eta_2)\,,\; (a_0 + \eta_1)\,,\; a_0\,,\; (a_0 \pm \eta_1)\,,$$
$$(a_0 \pm \eta_2)\,,\; \ldots\,,\; (a_0 \pm \eta_{n-1})\,,\; (a_0 \pm \eta_n)\,,$$

(*) On pourrait dire aussi « son terme *médian* », mais pas la *médiane* (qui sert généralement à désigner le terme médian d'une série rangée dans l'ordre croissant ou décroissant).

étant bien entendu que les η des termes équidistants du terme central ont toujours le même signe ou toujours des signes opposés.

Ainsi, pour prendre des exemples concrets, sont symétriques des séries comme les suivantes :

I	Série donnée (a)	20, 5, 10, 5, 20
	Écarts du terme central (η)	+ 10, — 5, 0, — 5, + 10
II	Série donnée (a)	4, 3, 5, 7, 6
	Écarts du terme central (η)	— 1, — 2, 0, + 2, + 1



GRAPHIQUE V.

Dans le cas où les différences entre chacun des deux termes équidistants et le terme central sont égales entre elles (cas I), les termes équidistants sont naturellement eux aussi égaux entre eux. Notre définition peut donc être formulée encore de la façon suivante : Une série est symétrique si les termes équidistants du terme central de la série sont égaux (cas I) ou si leurs écarts du terme central ont des valeurs absolues égales et des signes toujours opposés (cas II). Dans le premier cas nous dirons que la symétrie est *positive ;* dans le second, qu'elle est *négative*. Nous distinguerons ainsi des séries *positivement symétriques* (I) et des séries *négativement symétriques* (II).

§ 19. — D'un autre côté, on peut baser la définition de la symétrie sur les écarts *moyennaux* des termes de la série. Nous pouvons dire ainsi : *une série est symétrique si les écarts moyennaux des termes équidistants du terme central sont égaux ou ont des valeurs absolues égales et des signes toujours opposés.* Si nous désignons la moyenne de la série par a_m et les écarts moyennaux par d, cela veut dire qu'une série

$$a'_n\,,\, a'_{n-1}\,,\, \ldots\,,\, a'_2\,,\, a'_1\,,\, a_0\,,\, a''_1\,,\, a''_2\,,\, \ldots\,,\, a''_{n-1}\,,\, a''_n$$

est symétrique si :

$$d'_1\,(= a'_1 - a_m) = \pm\, d''_1\,(= \pm\,[a''_1 - a_m])\,,\, d'_2 = \pm\, d''_2\,,\ldots d'_n = \pm\, d''_n$$

(étant de nouveau entendu que d' et d'' ont soit toujours les mêmes signes, soit toujours des signes opposés). D'après cette définition une série symétrique peut être représentée encore comme suit :

$$(a_m + d_n)\,,\,(a_m + d_{n-1})\,,\,\ldots\,(a_m + d_2)\,,\,(a_m + d_1)\,,\,(a_m + d_0)\,,$$
$$(a_m \pm d_1)\,,\,(a_m \pm d_2)\,,\,\ldots\,(a_m \pm d_{n-1})\,,\,(a_m \pm d_n)$$

(même observation concernant les signes des d).

Cette définition coïncide en réalité avec notre définition première (du § 18).

En effet, en exprimant les termes généraux équidistants du terme central, d'une série (a'_n et a''_n) en fonction du terme central (a_0), nous avons :

$$a'_n = a_0 + \eta'_n\,;\; a''_n = a_0 + \eta''_n\,; \qquad \text{d'où :}$$
$$\eta'_n - \eta''_n = (a'_n - a_0) - (a''_n - a_0) = a'_n - a''_n\,. \qquad (1)$$

D'un autre côté, en exprimant les mêmes termes en fonction de la moyenne arithmétique de la série (a_m), nous avons :

$$a'_n = a_m + d'_n\;;\; a''_n = a_m + d''_n\;;\; \text{d'où :}$$
$$d'_n - d''_n = (a'_n - a_m) - (a''_n - a_m) = a'_n - a''_n \qquad (2)$$

Il résulte donc des deux égalités (1) et (2) que :

$$\eta'_n - \eta_n = d'_n - d''_n\,. \qquad [7]$$

Dans le cas où $\eta'_n = \eta''_n$ (symétrie positive d'après notre première définition), on trouve donc :

$$d'_n - d''_n = 0\,,\ \text{soit : } d'_n = d''_n\,,$$

c'est-à-dire que la série est positivement symétrique aussi d'après notre seconde définition. Dans le cas où $\eta'_n = -\eta''_n$, c'est-à-dire si les différences (η) entre les termes équidistants du terme central et ce dernier ont des valeurs absolues égales et des signes opposés, soit *dans le cas de symétrie négative* d'après notre première définition, *le terme central* (a_0) *est lui-même la moyenne arithmétique de la série.* En effet, si

$$a'_n = a_0 + \eta_n \quad \text{et} \quad a''_n = a_0 - \eta_n\,,$$

on a :

$$a'_n + a''_n = 2\,a_0 \;;\; \text{d'où : } a_0 = \frac{a'_n + a''_n}{2} = a_m \qquad [8]$$

Dans ce cas, notre définition première coïncide donc directement avec la seconde. Ce qu'il fallait démontrer.

Ainsi, dans les deux séries citées à titre d'exemple au § précédent, les écarts moyennaux sont les suivants :

		moyenne arithmétique
I. Symétrie positive :		
Série donnée (a)	20, 5, 10, 5, 20	12
Écarts moyennaux (d)	+ 8, — 7, — 2, — 7, + 8	
II. Symétrie négative :		
Série donnée (a)	4, 3, 5, 7, 6	5
Écarts moyennaux (d)	— 1, — 2, 0, + 2, + 1	

Les écarts moyennaux sont ici parfaitement symétriques tout comme les écarts par rapport au terme central (que nous appellerons écarts centraux).

§ 20. — Nous pouvons donner de la symétrie encore une troisième définition, basée sur les *écarts successifs* entre les termes de la série. En effet, si par écart successif d'un terme nous désignons la différence entre ce terme et le terme immédiatement précédent (en commençant de chaque côté par le terme central), nous pouvons dire qu'*une série est symétrique si les écarts successifs des termes équidistants du terme central sont égaux ou ont des valeurs absolues égales et des signes opposés*. Désignant les écarts successifs par e, nous pouvons donc dire qu'une série est symétrique si elle peut être exprimée sous la forme suivante (toujours la même observation concernant les signes) :

$$a'_n\,(= a'_{n-1} + e_n)\,,\; a'_{n-1}\,(= a'_{n-2} + e_{n-1})\,,\; \ldots\,,\; a'_2\,(= a'_1 + e_2)\,,$$

$$a'_1\,(= a_0 + e_1)\,,\; a_0\,,\; a''_1\,(= a_0 \pm e_1)\,,\; a''_2\,(= a''_1 \pm e_2)\,,\; \ldots\,.$$

$$a''_{n-1}\,(= a''_{n-2} \pm e_{n-1})\,,\; a''_n\,(= a''_{n-1} \pm e_n)\,.$$

Cette définition coïncide également avec les précédentes. En effet, dans toute série les écarts successifs des termes équidistants du terme central (a_0) sont égaux respectivement à :

$$e'_1 = a'_1 - a_0 = \eta'_1 \ ; \ e''_1 = a''_1 - a_0 = \eta''_1 \ ;$$

$$e'_2 = a'_2 - a'_1 = (a_0 + \eta'_2) - (a_0 + \eta'_1) = \eta'_2 - \eta'_1 \ ; e''_2 = \eta''_2 - \eta''_1 \ ;$$

$$e'_3 = a'_3 - a'_2 = (a_0 + \eta'_3) - (a_0 - \eta'_2) = \eta'_3 - \eta'_2 \ ; e''_3 = \eta''_3 - \eta''_2 \ ,$$

et ainsi de suite :

$$e'_n = \eta'_n - \eta'_{n-1} \ ; \ e''_n = \eta''_n - \eta''_{n-1} \ ;$$

Dans le cas de symétrie positive d'après notre première définition (§ 18), c'est-à-dire si $\eta'_1 = \eta''_1$, $\eta'_2 = \eta''_2$; .... $\eta'_n = \eta''_n$, on trouve donc que $e'_1 = e''_1$, $e'_2 = e''_2$, ... $e'_n = e''_n$, c'est-à-dire qu'il y a aussi parfaite symétrie positive d'après notre dernière définition. Dans le cas de symétrie négative d'après la première définition, puisque $\eta'_1 = -\eta''_1$, $\eta'_2 = -\eta''_2$, ... $\eta'_n = -\eta''_n$, on trouve que : $e'_1 = \eta'_1 = -\eta''_1 = -e''_1$, de même que $e'_2 = \eta'_2 = -\eta''_2 = -e''_2$, ... $e'_n = -e''_n$, c'est-à-dire qu'il y a symétrie négative aussi d'après notre 3me définition.

Pour reprendre les exemples cités dans les deux §§ précédents, on trouve pour les séries en question les écarts successifs que voici :

I. Symétrie positive :

Série donnée (*a*)	20, 5, 10, 5, 20
Écarts successifs (*e*)	+ 15, — 5, 0, — 5, + 15

II. Symétrie négative

Série donnée (*a*)	4, 3, 5, 7, 6
Écarts successifs (*e*)	+ 1, — 2, 0, + 2, — 1

§ 21. — Il va de soi que si la série a un nombre de termes pair, elle ne peut être symétrique que si les deux termes médians sont égaux ou accusent des écarts moyennaux de valeur absolue égale et de signes opposés. Comme centre de symétrie, on doit alors considérer un nombre qui est égal à la moyenne de ces termes et situé entre eux. Pour plus de généralité, au lieu de dire terme central, nous dirons dans la suite *centre de symétrie,* qui peut être soit un terme réel de la

série (le terme médian), soit la moyenne arithmétique des deux termes médians.

Dans une série parfaitement symétrique, le centre de symétrie doit coïncider soit avec la moyenne arithmétique de la série soit avec un maximum ou avec un minimum. Nous avons vu [8] qu'en cas de symétrie négative, le centre de symétrie coïncide avec la moyenne arithmétique ; dans le cas de symétrie positive (lorsque les termes équidistants du centre sont égaux entre eux), le centre de symétrie doit coïncider avec un maximum ou avec un minimum (notre courbe I du graphique V) (*).

Dans nos exemples, nous avons envisagé des séries *finies* ; dans ces séries, en cas de symétrie parfaite, le centre de symétrie doit évidemment coïncider avec le terme médian (ou la moyenne des deux termes médians) de la série. Quand il s'agit de séries *infinies* (naturellement, infinies des deux côtés ; autrement la série ne saurait être symétrique), il n'y a souvent pas de terme fixé a priori qui doive servir de centre de symétrie.

Les séries *périodiques typiques*, celles que nous avons particulièrement en vue dans la présente étude, ressemblent sous ce rapport à des séries infinies. Car une série périodique peut être commencée indifféremment par n'importe lequel de ses termes ; n'importe lequel de ses termes peut donc également être considéré comme terme médian de la série. A priori, chaque terme d'une série périodique typique doit donc être considéré comme pouvant éventuellement constituer le centre de symétrie de la série.

Des différences importantes existent par suite, au point de vue de la symétrie, entre les séries finies ordinaires, d'un côté, et les infinies et périodiques typiques, de l'autre. Une série finie ordinaire ne peut être symétrique que par rapport à un seul terme (ou point) de la série ; nous pouvons dire dans ce sens qu'elle ne peut être que *monosymétrique*. Par contre, une série périodique typique (ou infinie) peut être symétrique par rapport à plusieurs termes (ou points) de la série à la fois ; nous dirons donc que les séries périodiques (et infinies en général) peuvent être *polysymétriques*. Ainsi, par exemple, la série ordinaire

$$5, 8, 10, 8, 5$$

(*) Dans des séries polycéphales à nombre de sommets impair, le centre de symétrie positive peut être un maximum (ou minimum) secondaire qui peut coïncider avec la moyenne arithmétique de la série.

est symétrique par rapport au terme 10 et uniquement par rapport à lui, tandis que la série périodique correspondante (5 , 8 , 10 , 8) est symétrique à la fois par rapport au terme 10 et par rapport au terme 5, puisque la même périodicité peut être écrite aussi sous la forme (10, 8, 5, 8).

D'après le nombre des centres de symétrie d'une série périodique (ou infinie en général), on peut ainsi distinguer des séries monosymétriques, bis-symétriques, tri-symétriques, tétra-symétriques, etc.

Toute série périodique typique qui est symétrique positivement (courbe I du graphique V) *a au moins deux centres de symétrie*, séparés l'un de l'autre par la moitié de la série ; elle est donc forcément au moins *bisymétrique*.

Notons encore le cas particulier de la *sinusoïde :* elle est *tétrasymétrique ;* elle est en effet symétrique par rapport aux sinus de 0°, de 90°, de 180° et de 270° (c'est-à-dire qu'elle forme une courbe symétrique si on la dessine en prenant comme point médian n'importe lequel de ces quatre points).

On pourrait examiner encore le cas des séries, symétriques ou non, qui se composent de parties symétriques à l'intérieur, mais n'insistons pas.

§ 22. — Dans la réalité, cependant, la symétrie parfaite est extrêmement rare. Mais les séries réelles peuvent accuser une *symétrie partielle*, se rapprochant plus ou moins de la symétrie parfaite. Elles peuvent être, en d'autres termes, plus ou moins symétriques, plus ou moins asymétriques, selon que les écarts (les écarts à la moyenne, les écarts successifs ou les écarts au centre de symétrie) des termes équidistants du centre de symétrie ont des valeurs absolues plus ou moins égales et des signes plus ou moins systématiquement identiques ou opposés.

Au sujet de la symétrie partielle, nous ferons ici les deux observations que voici :

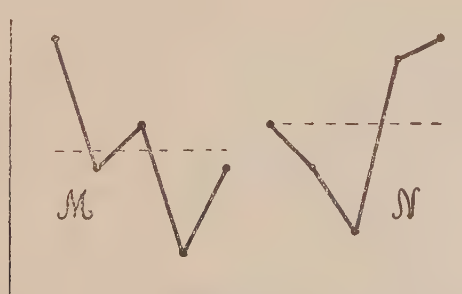
1° Dans le cas de symétrie partielle, le degré de symétrie (ou de dissymétrie) d'une série quelconque (donc finie ou infinie, périodique ou ordinaire) peut être envisagé par rapport à n'importe quel terme de la série pris comme centre de symétrie. Une série bien symétrique par rapport à tel de ses termes considéré comme centre de symétrie, peut l'être beaucoup moins ou, au contrarie, plus encore par rapport à tel autre. Ainsi, par exemple, la série

3, 5, 16, 20, 15, 6, 4, 1, 1,

est partiellement symétrique (ou dans une certaine mesure asymétrique) si l'on prend le terme médian 15 comme centre de symétrie ; elle se montre beaucoup plus symétrique si l'on prend comme centre de symétrie le terme 20, et apparaît très asymétrique par rapport aux termes 16 ou 6 pris comme centres de symétrie. Il va pourtant de soi que (sauf quelque but particulier), pour caractériser le degré de symétrie d'une série, on doit envisager sa symétrie vis-à-vis du terme où la symétrie se montre maxima (ou la dissymétrie minima).

2° Nous avons vu (§ § 10 et 20) qu'en cas de symétrie parfaite, les trois critères de symétrie (symétrie des écarts moyennaux, des écarts successifs et des écarts du centre) coïncident. Il n'en est pas forcément de même en cas de symétrie partielle. Une série relativement symétrique au point de vue de l'un de ces critères peut l'être beaucoup moins ou, au contraire, beaucoup plus d'après tel autre ; elle peut même accuser une certaine symétrie positive d'après tel critère et une notable symétrie négative d'après tel autre. Telles sont par exemple, les séries suivantes :

		Moyenne arithmétique
Série M	30, 15, 20, 5, 15	17
Écarts moyennaux (d) . . .	+ 13, — 2, + 3, — 12, — 2	
» successifs (e)	+ 15 — 5, — 15, + 10	
» du terme central (δ)	+ 10 — 5, — 15, — 5	
Série N	8, 6, 3, 11, 12	8
Écarts moyennaux (d) . . .	0, — 2, — 5, + 3, + 4	
» successifs (e)	+ 2, + 3, + 8, + 1	
» du terme central (δ)	+ 5, + 3, + 8, + 9	



GRAPHIQUE VI.

Nous examinerons plus loin le problème de la mensuration du degré de symétrie ; mais sans même procéder à des calculs de cet ordre, par la simple vue des courbes en question ou des chiffres relatifs aux écarts moyennaux, successifs et centraux, nous voyons que la série M n'accuse presque aucune symétrie des écarts à la moyenne ou des écarts au terme central, mais marque par contre une notable symétrie des écarts successifs ; la série N accuse à peine une certaine symétrie négative des écarts à la moyenne et, au contraire, une notable symétrie positive des écarts successifs et des écarts du terme central, et ainsi de suite. Quand il s'agit de symétrie partielle, c'est-à-dire dans la presque totalité des séries réelles, il faut donc spécifier le genre de symétrie dont on donne la mesure (*).

Notons enfin qu'il résulte de notre définition même de la symétrie que le degré d'asymétrie (ou de symétrie) d'une série dépend de la mesure de l'inégalité des écarts (moyennaux, successifs ou centraux) qu'accusent les termes équidistants du centre de la série ; autrement

---

(*) Dans ce travail nous ne nous occupons que de la symétrie *directe*. Mais il peut y avoir aussi des séries *indirectement* symétriques. Tel est notamment le cas de séries qui présentent des *fonctions symétriques*. Ainsi, par exemple, la série :

$$16,\ 8,\ 32,\ 128,\ 64$$

n'est pas directement symétrique (ou fortement asymétrique), mais elle peut être exprimée sous la forme :

$$2^4\ ,\ 2^3\ ,\ 2^5\ ,\ 2^7\ ,\ 2^6\ ;$$

elle présente donc une série de puissances de 2 parfaitement symétriques (série II, § 18). En prenant les logarithmes de ces mêmes termes, on obtient la série :

$$4 \lg 2\ ,\ 3 \lg 2\ ,\ 5 \lg 2\ ,\ 7 \lg 2\ ,\ 6 \lg 2$$

qui est aussi parfaitement symétrique.

Le cas examiné se rapporte aux fonctions négativement symétriques (dans le cas de fonctions positivement symétriques, la série donnée serait directement symétrique). Mais il se peut aussi que la série donnée ne soit pas symétrique et que certaines fonctions de ses termes présentent une parfaite symétrie positive. Tel est notamment le cas quand la dissymétrie de la série donnée provient des signes des termes. Ainsi, par exemple, la série :

$$10,\ 6,\ 20,\ 6,\ -10$$

n'est pas symétrique, mais les carrés de ces mêmes termes présentent une série parfaitement symétrique.

dit, il dépend de la valeur même des divers termes et de leur disposition dans la série. Telle série parfaitement symétrique cessera de l'être si ses termes changent de place. Telle autre série (par exemple : 6, 7, 11, 17, 12, 8, 4) ne sera jamais parfaitement symétrique à cause de la valeur de ses termes. Mais même dans ce dernier cas, la situation des termes n'est pas sans effet sur le degré de symétrie ou de dissymétrie de la série. Ainsi, la série que nous venons de citer est plus symétrique qu'une autre ayant les mêmes termes, mais autrement disposés (par exemple : 6, 7, 18, 12, 11, 8, 4). Le degré de symétrie d'une série est fonction de deux facteurs : de la valeur des termes et de leur disposition dans la série.

CHAPITRE IV.

La comparaison des séries périodiques et leur péréquation analitique.

§ 23. — Comme nous l'avons dit, l'étude des variations périodiques est basée sur la comparaison des diverses séries périodiques entre elles. Sous certains rapports, on peut comparer directement des séries périodiques en les mettant, elles ou les courbes qui les représentent, simplement les unes à côté des autres. Nous pouvons toujours comparer directement les séries pour ce qui concerne la durée de leur période, le nombre de leurs termes et de leurs sommets, la situation de leurs maxima et de leurs minima ainsi que la durée et la situation de chacune des phases dont la période se compose : la phase pendant laquelle le phénomène va en montant ou en baissant de même que la phase pendant laquelle il est supérieur ou inférieur à la moyenne de la période (phases de hausse et de baisse en comparaison avec le terme immédiatement précédent ou bien en comparaison avec la moyenne de la série). Dans certains cas, on peut établir encore d'autres comparaisons, surtout si les séries comparées présentent des variations qui gardent constamment le même sens et si l'une d'elles a constamment des variations plus fortes que l'autre ; on peut naturellement dire alors directement laquelle accuse un écartement moyennal ou successif plus fort ou plus faible. On peut parfois dire aussi directement si l'une des séries est plus ou moins concentrée dans ses écarts, plus ou moins précipitée, plus ou moins

symétrique que les autres. Inutile de dire que le plus fréquemment les séries périodiques qu'il s'agit de comparer ne varient pas toutes et toujours dans le même sens. Généralement aussi la hausse ou la baisse est à certains moments plus forte dans l'une des séries comparées et à certains autres moments dans une autre série, et ainsi de suite. Et même dans les cas relativement rares où la comparaison directe des séries nous permet de dire quelles variations sont plus fortes, concentrées, précipitées, etc., elle ne nous montre dans la réalité jamais dans quelle mesure elles le sont.

Il est donc nécessaire de recourir à des *expressions synthétiques* résumant sous une forme numérique les divers aspects des variations périodiques qui ne sont pas toujours directement comparables. Comparer ensuite ces expressions numériques revient à comparer les séries mêmes sous les rapports envisagés.

Quelles peuvent donc être ces expressions synthétiques ?

§ 24. — On pourrait croire que la solution en quelque sorte idéale de notre problème serait de trouver une formule analytique de péréquation des courbes périodiques, formule relativement simple et suffisamment souple pour s'adapter aux différentes périodicités existantes sans faire trop de violence à la réalité : on n'aurait ensuite qu'à comparer les valeurs concrètes de cette formule pour les diverses séries de variations données.

Les types de courbes à employer dans ce but peuvent être bien variés. Il n'est pas nécessaire même que la courbe théorique cherchée soit une courbe périodique puisqu'il s'agit de représenter géométriquement ou algébriquement la série périodique typique, donc une seule période. Toutefois, pour que la comparaison des formules analytiques concrètes soit assez simple, il serait à désirer que les courbes qu'elles expriment appartiennent au même type.

De fait, c'est la *sinusoïde* qui a été mise en avant dans le but de la péréquation des variations périodiques. La sinusoïde, qui est une courbe périodique et qui a pour formule fondamentale : $y = a + b \sin x$ paraît en effet se prêter assez bien à ce but.

D'abord, elle est relativement facile à calculer : en mettant a égal à la moyenne de la période ($= v_m$), la variable x égale respectivement à $\frac{1}{n} 360°$, $\frac{2}{n} 360°$, $\frac{3}{n} 360°, \ldots \frac{n-1}{n} 360°$ et $360°$ (n étant le nombre des termes de la série périodique) et b égal à $\frac{\Sigma (v \sin x)}{\Sigma \sin^2 x}$,

on trouve aisément les y qui forment la série théorique dont la série périodique donnée est plus ou moins rapprochée, série théorique qui varie entre $(a + b)$ au maximum et $(a - b)$ au minimum et où les $b \sin x$ expriment ce que nous avons appelé les écarts moyennaux de la série périodique.

D'un autre côté, dans cette formule a représente ce qui est indépendant des variations périodiques ; les $\sin x$ sont déterminés par le nombre des termes de la série (n) et par la place de chaque terme dans la série ; pour des séries homonomes (n constant), les termes occupant la même place ont donc les mêmes $\sin x$; le diverses formules concrètes des variations périodiques ne diffèrent ainsi que par la valeur du paramètre b qui marque l'intensité de ces variations (ou leur écartement moyennal). La comparaison des b ou, en d'autres termes, les rapports $\frac{b'}{b''}$ nous donnent ainsi l'intensité relative des séries de variations périodiques considérées.

§ 25. — En passant, signalons deux propriétés de la sinusoïde que nous croyons inconnues jusqu'ici et qui ne sont pas sans intérêt pour notre problème.

1° On peut démontrer que *si l'on divise l'axe des abscisses de la sinusoïde* $[y = \sin x]$, *correspondant aux* 360° *de la circonférence du cercle, en* n *parties égales* (n *étant* > 2), *la somme des carrés des ordonnées de la sinusoïde menées à chaque division est égale à* $\frac{n}{2}$. Cette proposition peut être formulée encore de la façon suivante :

Si l'on divise la circonférence du cercle en n *parties égales* (n *étant* > 2), *la somme des carrées des sinus des divers arcs* $\left(\frac{1}{n}360°, \frac{2}{n}360°, \frac{3}{n}360°, \ldots \frac{n-1}{n}360°, 360°\right)$ *ainsi obtenus est égale à* $\frac{n}{2}$, soit :

$$\sin^2\left(\frac{1}{n}360°\right) + \sin^2\left(\frac{2}{n}360°\right) + \sin^2\left(\frac{3}{n}360°\right) + \ldots + \\ + \sin^2\left(\frac{n-1}{n}360°\right) + \sin^2 360° = \frac{n}{2}. \qquad [9]$$

Ainsi, la somme des carrés des 12 sinus de la circonférence correspondant aux arcs de 30°, 60°, 90°,. . . 330° et 360° (ou des 12 ordonnées de la sinusoïde [$y = \sin x$] menées à chaque $\frac{1}{12}$ de son axe des abcisses) sera égale à 6 ; la somme des carrés des 24 ordonnées de la sinusoïde menées à des intervalles égaux du point 0 sera = 12 ; la somme des carrés de 7 ordonnées menées à des intervalles égaux du point 0 sera = 3,5, et ainsi de suite.

Cette propriété simplifie encore beaucoup le calcul du paramètre de la péréquation b en nous dispensant de calculer la somme des carrés des $\sin x$ ($\Sigma \sin^2 x$) qui forme le dénominateur de la fraction qui nous donne la valeur de b. Nous pouvons donc, à la place de $b = \frac{\Sigma (v \sin x)}{\Sigma \sin^2 x}$, prendre simplement :

$$b = \frac{\Sigma (v \sin x)}{n : 2}, \qquad [10]$$

soit : pour une série à 12 termes, $b_{12} = \frac{\Sigma (v \sin x)}{6}$

» » » 24 » $b_{24} = \frac{\Sigma (v \sin x)}{12}$

» » » 7 » $b_7 = \frac{\Sigma (v \sin x)}{3{,}5}$.

2° *Dans la sinusoïde, l'écart moyennal quadratique moyen* ($\sigma_{\sin}$) *est égal à* $b \sqrt{\frac{1}{2}}$ *ou, en mettant* $b = 1$, *simplement à* $\sqrt{\frac{1}{2}} = 0{,}707$.

En effet, les écarts moyennaux de la sinusoïde [$y = a + b \sin x$] n'étant autre chose que les ordonnées de la sinusoïde [$y' = b \sin x$], nous avons :

$$\sigma_{\sin} = \sqrt{\frac{\Sigma (b \sin x)^2}{n}} = \sqrt{\frac{b^2 \Sigma \sin^2 x}{n}} = b \sqrt{\frac{\Sigma \sin^2 x}{n}} ;$$

mais d'après [9], $\Sigma \sin^2 x = \frac{n}{2}$; il s'ensuit donc que

$$\sigma_{\sin} = b \sqrt{\frac{n}{2\,n}} = b \sqrt{\frac{1}{2}} \qquad [11]$$

L'écart moyennal quadratique moyen est ainsi pour la sinusoïde indépendant du nombre des intervalles qui la divisent. Il s'ensuit, de notre dernière formule que pour la sinusoïde :

$$\frac{b}{\sigma_{\text{sin}}} = \sqrt{2} = 1{,}414\,, \qquad [12]$$

c'est-à-dire que *dans la sinusoïde, le rapport du paramètre* (b) *à l'écart moyennal quadratique moyen* (σ) *est constant et égal à* $\sqrt{2}$.

Mais b n'est que la moitié de l'amplitude des variations de la sinusoïde, amplitude que nous désignons d'habitude par α. La dernière expression peut donc être écrite encore comme suit :

$$\frac{\alpha}{2\sigma_{\text{sin}}} = 1{,}414 \text{ (*)} \qquad [13]$$

L'expression $\frac{\alpha}{2\sigma}$ $\left(\text{ou } \frac{\alpha}{\sigma}\right)$ présente une intéressante combinaison des deux notions fondamentales de toute variation : de la notion des extrêmes (α) et de celle de la moyenne (σ) ; pour la sinusoïde, comme nous le voyons ici, elle est constante. Elle peut ainsi servir aussi de point de repère pour l'orientation dans certains caractères des variations périodiques réelles.

§ 26. — Mais la péréquation par la sinusoïde constitue-t-elle une solution véritable de notre problème ?

Si le but de la péréquation est de fixer la valeur globale approximative d'un phénomène à un moment donné de sa variation périodique, on peut réellement admettre que la péréquation par la sinusoïde réalise souvent ce but (avec une approximation d'ailleurs très variable) ; tel est notamment le cas si $\frac{b}{a}$ est une petite fraction ou, si l'on préfère, lorsque la partie variable (l'écartement moyennal, b) est relativement peu importante par rapport à la partie constante (la moyenne de la période, a) de la grandeur considérée (**). Mais si

(*) On pourrait simplifier encore cette formule en écrivant : $\frac{\alpha}{\sigma_{\text{sin}}} = 2{,}828$. Mais α étant la somme des *deux* écarts extrêmes, son rapport au *double* de l'écart quadratique moyen (2σ) nous paraît plus logique.

(**) L'exemple de la coïncidence la plus parfaite que je connaisse entre la série des variations empiriques et leur péréquation par la sinusoïde est

le but qu'on se propose est de dégager les caractères des variations proprement dites, toute péréquation, qui a nécessairement un effet niveleur, et en particulier la péréquation par la sinusoïde, doit généralement se moutrer un prodédé très déféctueux.

En effet, sous sa forme fondamentale et la seule suffisamment simple pour entrer ici en ligne de compte, la sinusoïde est une courbe symétrique, ondulatoire, relativement peu concentrée, ayant un seul sommet qui se trouve à égale distance des deux minima successifs. En substituant aux différentes séries périodiques réelles leur péréquation par la sinusoïde, on leur enlève, pour la plus grande partie, leurs caractères propres et on leur communique les propriétés mentionnées de la sinusoïde. Leur degré de concentration, de précipitation, de dissymétrie, le nombre de leurs sommets, — tout cela disparaît avec cette péréquation. Même de l'écartement moyennal des séries périodiques, la péréquation par la sinusoïde ne donne une idée suffisamment rapprochée de la réalité que si par ses autres caractères (symétrie, concentration, gradation, monocéphalie), la série empirique se rapproche de la sinusoïde.

Dans la pratique, il faut donc d'abord voir si la série donnée est monocéphale, relativement graduelle (ondulatoire) et symétrique (ce qui s'aperçoit directement). Si tel est le cas, la péréquation par la sinusoïde présente naturellement un certain intérêt (*). Dans

---

celui des variations de la température du corps humain prise toutes les deux heures d'après VIAULT et JOLYET, cité par le Professeur NICEFORO dans son traité *La méthode statistique et ses applications* (Paris 1925, p. 304-305). L'écart moyen entre la série empirique et la série ajustée par la sinusoïde n'est dans cet exemple que de 0,4 % de la température moyenne de la journée. Mais dans cet exemple la coïncidence est due beaucoup moins au procédé de péréquation qu'au fait que les fluctuations horaires de la température (en moyenne 0°,3) ne portent même pas sur $\frac{1}{100}$ de la température moyenne de la journée (36°,75). Si au lieu de rapporter l'écart moyen entre les deux séries à la *valeur globale* moyenne, on le rapporte à l'*écart* horaire moyen de la température, c'est-à-dire si l'inexactitude commise par la péréquation est comparée avec l'élément variable de la température (dont il s'agit précisément de rendre compte !), on trouve qu'elle n'en constitue pas moins de 50 p. 100.

(*) Nous avons nous-même montré que dans certains cas les fluctuations saisonnières (typiques) du chômage peuvent être représentées par une sinusoïde. Voyez notre étude déjà citée *Les fluctuations saisonnières du chômage dans l'industrie du bâtiment*, notamment pour l'industrie du bâtiment au Royaume-Uni ainsi que pour l'industrie des constructions et pour l'ensemble de l'industrie en Italie (pp. 36, 42 et 73-74 du tirage à part. Voyez aussi plus loin, §§ 74-76).

le cas contraire, cette péréquation est un trompe-l'œil bien plus qu'une image vraie de la réalité, un trompe-l'œil particulièrement grave par suite de son apparence mathématique.

Or, il va de soi que toutes les variations périodiques sont loin d'avoir une allure très sinusoïdale, que pareilles variations sont dans la réalité plutôt peu fréquentes. Nous pensons qu'il doit en être de même, peut-être encore *a fortiori*, de la péréquation par une courbe de quelque autre type.

Ce qu'il nous faut pour pouvoir comparer les différentes périodicités entre elles afin de dégager les caractères propres à chacune d'elles, ce n'est donc pas, semble-t-il, quelque formule générale de péréquation, mais un système d'indices numériques dont chacun se rapporterait à un certain caractère déterminé des variations périodiques, résumant ainsi un certain aspect de la série envisagée. Les formules de péréquation semblent plutôt devoir jouer ici un rôle très secondaire, ne pouvant dans la réalité s'appliquer raisonnablement qu'à des périodicités fort peu nombreuses.

CHAPITRE V.

Grandeurs numériques variant avec certains caractères des séries périodiques.

§ 27. — Nous avons constaté plus haut (§ 23) que sous certains rapports les séries périodiques peuvent être comparées directement (par exemple, pour le nombre des sommets, la situation des maxima et des minima, la durée de la période, etc.). Nous avons vu également (§ 10) que si l'*intensité des fluctuations* des diverses séries périodiques ne peut généralement pas être comparée directement, il existe pourtant des expressions numériques qui varient avec cette intensité et qui, elles, pourraient éventuellement être comparées ; ce sont notamment : l'amplitude des variations (α et α'), l'écart moyennal moyen et l'écart successif moyen. Après ce que nous avons vu plus haut (§§ 10-14 et 25), il est inutile d'insister encore sur l'importance des valeurs extrêmes des séries périodiques typiques. Arrêtons-nous cependant un moment sur les deux autres points.

L'écart moyennal moyen d'une série peut être pris sous la forme d'une moyenne arithmétique $\left(d_m = \frac{\Sigma d}{n}\right)$ ou sous celle d'une moyenne

quadratique $\left(\sigma = \sqrt{\frac{\Sigma d^2}{n}}\right)$. Sa signification dans les deux cas n'est pas la même. Pour plus de simplicité, prenons le cas de séries homonomes (n étant constant). L'écart moyennal *arithmétique* moyen varie dans ce cas uniquement avec la somme des valeurs absolues des écarts moyennaux (Σd) et demeure entièrement *insensible à l'égalité ou à l'inégalité de ces écarts entre eux*, c'est-à-dire à la *concentration* ou à l'extension (nivellement) des variations. Ainsi, par exemple, dans les deux séries périodiques A et B (§ 15), dont l'une (A) est très « concentrée » et l'autre (B) est au contraire très « étendue » ou « nivelée », l'écart arithmétique moyen demeure le même ($d_m = 1$). Par contre, l'écart *quadratique* moyen se trouve fortement *influencé aussi par l'inégalité* (*ou l'égalité*) *de la répartition de la somme des écarts moyennaux* entre les différents termes de la série, donc influencé aussi par la *concentration* ou l'extension (nivellement) des fluctuations. Ainsi dans le même exemple, σ est très différent pour A et pour B. Pour A (périodicité très concentrée, répartition très inégale de l'écartement moyennal), l'écart quadratique moyen est égal à :

$$\sigma_A = \sqrt{\frac{6^2 + 6^2}{12}} = \sqrt{\frac{72}{12}} = \sqrt{6} = 2{,}45$$

tandis que pour B (extension maxima, répartition parfaitement égale de l'écartement moyennal), il est égal à :

$$\sigma_B = \sqrt{\frac{12}{12}} = 1\,.$$

Dans la suite (§§ 45-46) nous verrons de plus près comment σ varie avec la concentration et l'extension (nivellement) des variations.

De plus, de notre définition même de la concentration et du nivellement (extension) des variations (§ 15), il résulte directement qu'avec la concentration ou le nivellement varie *l'écart moyennal moyen du second ordre*. Et effectivement, cet écart, qui dans le cas de concentration nulle (ou de nivellement maximum), comme nous l'avons vu (§ 15), devient égal à 0 (cas de la série B), est égal à 1,67 dans le cas d'une série très concentrée comme celui de la serie A.

§ 28. — *L'écart successif moyen* peut lui aussi être envisagé comme écart arithmétique moyen $\left(e_m = \frac{\Sigma e}{n}\right)$ et comme écart quadratique moyen $\left(\text{que nous désignerons par } \varsigma = \sqrt{\frac{\Sigma e^2}{n}}\right)$.

L'écart successif *arithmétique* moyen, comme nous l'avons vu plus haut (§ 13, 2° et § 15, 3°), dépend uniquement des amplitudes des variations des séries et du nombre de leurs termes. Pour des séries homonomes, il n'ajoute donc rien à ce que nous donnent les amplitudes des variations (α et α'). Par contre, l'écart successif *quadratique* moyen (ς), en cas de séries *homocéphales*, présente une parfaite analogie avec l'écart moyennal quadratique moyen, et se trouve fortement *influencé aussi par l'égalité ou l'inégalité des écarts successifs* entre eux, c'est-à-dire par le caractère plus ou moins *graduel* (ondulatoire) ou *précipité* de la série. Ainsi, si nous reprenons l'exemple de deux séries monocéphales et homonomes X et Y (§ 17), ayant la même amplitude ($\alpha = 6$) et par suite aussi le même écart successif arithmétique moyen ($e_m = 1$), mais dont l'une (X) a une allure très précipitée, ses écarts successifs étant très inégaux, et dont l'autre (Y) est au contraire parfaitement graduelle, ondulatoire, ses écarts successifs étant tous égaux entre eux, on trouve :

pour X (variations très précipitées) :

$$\varsigma_X = \sqrt{\frac{3^2 + 6^2 + 3^2}{12}} = \sqrt{\frac{54}{12}} = 2{,}12$$

pour Y (variations parfaitement graduelles) :

$$\varsigma_Y = \sqrt{\frac{12}{12}} = 1 .$$

L'écart successif quadratique moyen (ς) varie donc non seulement avec la somme des écarts successifs (soit avec les amplitudes et le nombre des sommets de la série), mais aussi avec l'allure plus ou moins précipitée ou ondulatoire de la série. Plus loin (§ 54), nous verrons aussi comment ς varie avec cette allure de la série.

Toujours dans le cas de séries homocéphales, l'écartement successif des termes de la série présente encore une analogie avec leur écartement moyennal. En effet, de notre définition de la précipita-

tion et de la gradualité (ondulation) des variations (§ 17), il s'ensuit qu'avec la précipitation (ou la gradualité) varie l'*écart successif moyen du second ordre*. Ceci se vérifie en particulier également dans l'exemple des deux séries monocéphales X et Y : dans la série Y, qui présente une gradualité parfaite (précipitation minima), l'écart successif moyen du second ordre est égal à 0, tandis que dans la série X ayant la même amplitude, mais dont les variations sont très précipitées, ce même écart est égal à 2.

§ 29. — On s'est souvent occupé de la *symétrie* et de la *dissymétrie* de séries statistiques. Mais pour autant que nous le sachions, il s'agissait presque toujours de séries soumises à la loi de la courbe « normale » des écarts accidentels. C'est aussi la symétrie de cette courbe théorique qu'on avait en vue parlant du degré de symétrie ou d'asymétrie des séries (ou courbes) réelles. Et l'on fondait la notion même de la symétrie sur certaines propriétés de la courbe « normale ». Ainsi, comme dans cette courbe la moyenne arithmétique, la médiane et le mode coïncident, on a le plus souvent basé l'indice de dissymétrie des séries réelles sur la différence entre ces trois grandeurs. D'un autre côté, dans la courbe des écarts accidentels, la fréquence des écarts positifs est égale à celle des écarts négatifs des mêmes dimensions, c'est-à-dire la somme algébrique des cubes des écarts moyennaux y est égale à 0 ; on a par suite, pris aussi cette somme ou, plus exactement, la racine cubique de cette somme divisée par le nombre des termes, comme indice (avec certaines modifications, même comme « cœfficient ») de dissymétrie.

Introduits par des statisticiens éminents, ces indices et « cœfficients » passèrent ensuite dans toute une série de manuels et de traités de statistique, comme si l'on avait là . . . vraiment des indices et des cœfficients de dissymétrie.

Au fond cependat, dans les écarts accidentels, il s'agit d'une certaine espèce de symétrie et non de la symétrie en général. Pour se rendre clairement compte de quelle espèce de symétrie il s'agit ici, on n'a qu'à laisser pour un moment de côté la courbe de Gauss, qui donne du phénomène une image qu'on pourrait appeler en un certain sens cumulative, et se reporter à *l'ogive de Galton*. Cette ogive, dans le cas d'écarts purement accidentels, est parfaitement symétrique : son terme médian (centre de symétrie) coïncide avec la moyenne arithmétique de la série, et les termes équidistants de ce centre de symétrie accusent des écarts de la moyenne qui sont égaux entre

eux et qui ont toujours des signes opposés. C'est donc une *symétrie négative* d'après notre terminologie (§ 19).

C'est en effet dans cette espèce de symétrie, et dans celle-ci seulement, que la médiane coïncide nécessairement avec la moyenne arithmétique. Par contre, aucune nécessité d'une telle coïncidence n'existe pour les séries qui sont symétriques *positivement*. Bien plus, dans les séries symétriques de cette dernière espèce, une telle coïncidence, toute fortuite, ne peut se rencontrer que très rarement.

Il en est de même de la somme algébrique des cubes des écarts moyennaux : elle est forcément égale à 0 dans des séries négativement symétriques, mais nullement dans celles qui sont symétriques positivement. Quant à la notion du « mode », elle ne se justifie guère en dehors de variations accusant des écarts accidentels ; elle n'a qu'une signification fort limitée quand il s'agit de séries périodiques typiques (et de séries typiques en général) ; elle n'a en tout cas aucun titre pour servir, dans des séries typiques, de base à un étalon de symétrie et d'asymétrie.

Il serait peut-être curieux de noter en passant que les *séries du triangle de Pascal*, ou les séries dont les termes s'exprimeraient par les *coefficients du binôme* développé, *ne seraient pas symétriques*, d'après les indices de symétrie (et d'asymétrie) basés soit sur la différence entre médiane et moyenne, soit sur la racine cubique des cubes des écarts moyennaux. Ainsi, par, exemple, on trouve pour la série des coefficients du binôme pris à la puissance

	Coefficients du binôme	Moyenne arithmétique	Médiane	$\sqrt[3]{\frac{\Sigma d^3 (*)}{n}}$
4ᵉ	1, 4, 6, 4, 1	3, 2	4	0,7
6ᵉ	1, 6, 15, 20, 15, 6, 1	9,14	6	5,1
7ᵉ	1, 7, 21, 35, 35, 21, 7, 1. . .	16	14	9,0
9ᵉ	1, 9, 36, 84, 126, 126, 84, 36, 9, 1	51, 2	36	36,8

et ainsi de suite. Toutes ces séries parfaitement symétriques auraient ainsi d'importants indices d'asymétrie ; ces indices se montreraient même d'autant plus forts que la puissance du binôme est plus élevée (ou que les termes de la série s'écartent davantage de sa moyenne).

Et rien d'étonnant ; car les coefficients du binôme présentent une

(*) Il s'agit ici de la somme *algébrique*.

série de nombres qui sont symétriques *positivement*. (En effet, dans la courbe binômiale, tandis que les abcisses rangées d'après leur ordre de grandeur présentent une série symétrique négativement, les ordonnées, elles, forment une série symétrique positivement).

Notons encore que ces indices d'asymétrie, basés sur la moyenne arithmétique de la série et les écarts moyennaux, sont insensibles à la symétrie (ou dissymétrie) des *écarts successifs* ou à celle des écarts *centraux*. Or, comme nous l'avons vu plus haut (§ 22, 2°), en cas de symétrie partielle (donc dans la presque totalité des cas réels), le degré de symétrie (ou d'asymétrie) des écarts moyennaux peut être bien différent de celui des écarts succesifs ou centraux pour lequel les indices en question demeurent entièrement muets.

Enfin, ces prétendus indices et coefficients d'asymétrie ont encore un autre grave défaut qui rend leur usage fallacieux même en tant qu'indices de la seule symétrie négative et moyennale : c'est qu'ils sont *insensibles à la situation des termes dans la série*. Or, nous avons vu (§ 22) que la symétrie d'une série est fonction à la fois de la valeur de ses termes et de leur disposition dans la série. Mais avec les indices en question, des séries ayant les mêmes termes, mais autrement disposés, c'est-à-dire des séries pouvant être très différentes au point de vue de la symétrie, accuseront des indices de symétrie (ou d'asymétrie) identiques. Ainsi, par exemple, des séries bien asymétriques comme :

5, 8, 3, 10, 12, 15, 17
17, 8, 15, 10, 3, 5, 12
12, 17, 15, 10, 3, 5, 8
ou même 17, 3, 10, 15, 12, 8, 5, etc.

auraient exactement le même indice d'asymétrie $\left(\sqrt{\frac{\Sigma\, d^3}{7}} = 0\right)$ que la série parfaitement symétrique négativement :

8, 5, 3, 10, 17, 15, 12.

En général, il nous paraît de toute évidence qu'une grandeur qui n'est pas sensible à la disposition des termes ne peut pas, par définition, servir d'indice de symétrie ou d'asymétrie.

§ 30. — Pour trouver ce qui varie effectivement avec le degré de symétrie ou de dissymétrie, on n'a qu'à remonter à la définition

de la dissymétrie. Une série est symétrique (§§ 18-20) si les termes équidistants du centre de symétrie accusent des écarts (moyennaux, successifs ou centraux) soit égaux (symétrie positive) soit de valeur égale et de signes opposés (symétrie négative). Lorsque cette égalité n'existe pas, la série est plus ou moins asymétrique. Et plus ces écarts des termes équidistants s'éloignent de l'égalité — pour la symétrie positive — ou de la valeur opposée (valeurs absolues égales et signes opposés) — pour la symétrie négative —, et plus aussi la série est asymétrique. En cas de symétrie parfaite positive, les *différences* des écarts des termes équidistants sont toutes égales à zéro ; et plus ces différences sont grandes (comme valeurs absolues (*)), et plus la série devient asymétrique. En cas de parfaite symétrie négative, la *somme algébrique* des écarts des termes équidistants est pour tous les termes égale à zéro ; et plus les valeurs absolues de cette somme sont grandes, et plus aussi la série est asymétrique. *Ce sont les différences* (pour la symétrie positive) *ou les sommes* (symétrie négative) *des écarts des termes équidistants qui varient avec le degré de symétrie et d'asymétrie de la série.* Un indice de symétrie ou d'asymétrie doit être basé sur la comparaison des écarts (moyennaux, successifs ou centraux) des termes équidistants.

CHAPITRE VI.

Indices de certains caractères des séries périodiques.

§ 31. — Chacune des grandeurs que nous avons retenues dans le précédent chapitre varie avec tel ou tel autre caractère des séries, mais pas uniquement avec lui. Elles sont influencées encore par d'autres facteurs. Pour qu'elles puissent servir d'indices d'intensité des caractères envisagés, il faut donc d'abord *éliminer* l'action des autres facteurs. Très souvent, nous atteingnons ce but en substituant des chiffres relatifs aux chiffres absolus ou, plus généralement, en divisant les

---

(*) Nous disons *comme valeurs absolues*, car le degré d'asymétrie est naturellement le même quand nous avons, par exemple, un écart de 7 à gauche et de 5 à droite (différence = 2) ou 5 à gauche et 7 à droite (différence = — 2).

chiffres réels (qui sont des produits combinés de plusieurs facteurs) par les facteurs qu'il s'agit d'éliminer (*).

Au fond, la notion même de la *moyenne* n'est qu'une application particulière de cette méthode générale : la *somme* des valeurs d'un certain groupe de phénomènes est fonction de la nature de ces phénomènes et de leur nombre ; en divisant cette somme par leur nombre, on obtient une grandeur (la moyenne arithmétique) que nous considérons comme indépendante du *nombre* des phénomènes envisagés et que l'on peut, par exemple, comparer avec les données (moyennes) d'autres groupes comprenant des phénomènes analogues en nombres différents. En particulier, la somme des écarts d'une série est fonction des variations des termes de la série et du nombre de ces termes ; *l'écart moyen* nous donne une idée de la variation de ses termes abstraction faite de leur nombre ; il nous permet, entre autre, de comparer les variations de séries ayant des nombres de termes différents. L'écart moyen lui-même est encore fonction de la variabilité des phénomènes envisagés et de leurs dimensions ; un écart moyen de 1 mm., par exemple, ne signifie évidemment pas une variabilité identique de la longueur pour les noix et pour les feuilles d'un noyer ; en divisant l'écart moyen des termes d'une série par la dimension moyenne de ces termes, nous obtenons un écart moyen relatif (par exemple, en % %) qui présente à nos yeux la variabilité des phénomènes indépendamment de leurs dimensions et qui nous paraît comparable à la variabilité d'autres phénomènes ayant des dimensions différentes (**).

Là où plus de deux facteurs entrent en ligne de compte, on peut les éliminer *un à un* de façon à ne laisser subsister (pratiquement) que le facteur cherché ; c'est ce que nous venons d'indiquer dans l'exemple de la variabilité en éliminant d'abord l'action du *nombre* des faits observés par le procédé de la moyenne $\left[\text{en prenant } v_m = \frac{\Sigma v}{n} \text{ et } d_m = \frac{\Sigma d}{n}\right]$ et en éliminant ensuite l'influence des *dimensions* des

---

(*) Voyez dans cet ordre d'idées l'étude riche en suggestions du Prof. GINI : *Quelques considérations au sujet de la construction des nombres indices et des questions analogues* dans *Metron* 1924.

(**) Nous disons « nous paraît », « nous considérons comme », etc., car ce procédé repose évidemment sur le postulat de la proportionnalité qui dans bien des cas fait défaut. Là où cette proportionnalité n'est pas admissible, ce procédé doit être remplacé par d'autres qui correspondent mieux à la nature des faits envisagés.

phénomènes par la division de l'écart moyen par la moyenne de la série $\left[\frac{d_m}{v_m}\right]$. Mais on peut aussi éliminer *d'emblée* l'action de *plusieurs* facteurs en divisant le chiffre donné par un autre qui présente l'action combinée des facteurs qui doivent être éliminés ; ainsi dans notre exemple, on pourrait directement diviser la somme des écarts par la somme des termes (longueurs de feuilles, noix, etc) $\left[\frac{\Sigma\, d}{\Sigma\, v}\right]$ pour obtenir l'écart moyen exprimé en % de la dimension des choses observées (éliminant ainsi à la fois et l'influence du nombre et celle des dimensions des phénomènes considérés).

C'est de la même façon que nous procèderons pour dégager les indices des caractères des séries périodiques : des chiffres bruts donnés, présentant l'effet combiné de ces caractères et d'autres facteurs, nous passerons à des chiffres relatifs en divisant les chiffres donnés par les facteurs qu'on doit éliminer.

Il va pourtant de soi que l'élimination des différents facteurs qui entrent en jeu est très souvent incomplète, soit qu'il nous est impossible de les éliminer tous, soit qu'intentionnellement nous laissons subsister l'action combinée de plusieurs d'entre eux. Dans ce dernier cas, nous pouvons, pour le même phénomène, calculer plusieurs chiffres relatifs selon les facteurs que nous éliminons et ceux que nous laissons, au contraire, subsister (*).

§ 32. — *L'amplitude* brute des variations d'une série (α) caractérise peu ces variations et, par suite, se prête relativement peu à la comparaison de différentes séries entre elles. Il faut avant tout tenir compte des dimensions (de la valeur) des termes de la série. En divisant l'amplitude des variations (α) par la moyenne arithmétique de la série (v_m), nous obtenons l'expression :

$$\alpha_v = \frac{\alpha}{v_m} \qquad [14]$$

qui élimine l'influence des dimensions des termes de la série , qui permet la comparaison de séries très différentes par la grandeur de leurs

(*) Aux différences formelles entre chiffres absolus et relatifs examinées plus haut (§ 7), répondent ainsi, au fond, des différences d'élimination.

termes et qui peut par suite être considérée comme un indice de l'amplitude des variations des séries.

On peut aussi diviser l'amplitude des variations (α) par l'écart moyennal moyen de la série. Ce dernier étant l'effet de l'action combinée des dimensions des termes de la série et de leur variabilité moyennale, on élimine par cette division à la fois ces deux facteurs. Comme écart moyennal moyen, on peut prendre soit la moyenne arithmétique (d_m) soit la moyenne quadratique (σ) des écarts moyennaux. On obtient ainsi les indices

$$\alpha_d = \frac{\alpha}{d_m} \ (*) \qquad [15]$$

$$\text{et } \alpha_\sigma = \frac{\alpha}{\sigma} \qquad [16]$$

qui donnent une idée de l'amplitude des variations des séries indépendamment des dimensions de leurs termes et de leur variabilité moyennale.

Quant à la signification propre de l'un et de l'autre de ces deux indices, il faut se rappeler ce que nous avons vu plus haut (§ 27) au sujet de la différence entre d_m et σ : tandis que, pour le même nombre de termes, l'écart arithmétique moyen (d_m) varie seulement avec la somme des écarts moyennaux, l'écart quadratique moyen (σ) représente l'action combinée de la somme de ces écarts et de leur concentration ou extension (nivellement). Par suite, $\frac{\alpha}{d_m}\left(\text{ou } \frac{\alpha}{2\, d_m}\right)$ nous donne l'amplitude relative des variations après élimination de la valeur des termes et de leur écartement moyennal tandis que $\frac{\alpha}{\sigma}\left(\text{ou } \frac{\alpha}{2\,\sigma}\right)$ élimine encore la différence de concentration de cet écartement.

Pour la sinusoïde, comme nous l'avons vu plus haut [13], le rapport $\frac{\alpha}{2\,\sigma}$ est constant et égal à 1,414. La valeur concrète de ce rap-

(*) α étant la somme des *deux* écarts moyennaux maxima, on ferait mieux, comme nous l'avons déjà remarqué, de prendre comme dénominateur le double de l'écart moyen, donc : $\alpha_d = \frac{\alpha}{2\, d_m}$ et $\alpha_\sigma = \frac{\alpha}{2\,\sigma}$ $\left(\text{on pourrait mettre de même : } \alpha_v = \frac{\alpha}{2\, v_m}\right)$.

port pour une série périodique donnée constitue donc, en outre, un des éléments permettant de juger dans quelle mesure cette série se rapproche de la sinusoïde.

La division de α par l'écart *successif* arithmétique moyen (e_m), quand il s'agit de comparer des séries périodiques monocéphales, ne présente aucun intérêt. Car nous avons déjà vu [3] que pour cette espèce de séries, le rapport $\frac{\alpha}{e_m} = \frac{n}{2}$, c'est-à-dire qu'il dépend uniquement du nombre des termes (n) de la série et qu'il est constant pour des séries homonomes. Quant au rapport $\frac{\alpha}{\varsigma}$ (l'amplitude divisée par l'écart successif quadratique moyen), nous le retrouverons encore plus loin (§ 35).

§ 33. — *L'écart moyennal arithmétique moyen* des termes d'une série (d_m) est fonction de la variabilité moyennale des phénomènes considérés et de leurs dimensions. En divisant cet écart par la moyenne des termes de la série (v_m), nous obtenons l'expression :

$$d_{m\,(v)} = \frac{d_m}{v_m} \tag{17}$$

qui nous donne un indice de la variabilité moyennale de la série considérée.

D'habitude, on prend à cet usage l'écart moyennal *quadratique* moyen (σ) considéré comme l'écart type (*standard deviation*) par excellence. Mais lorsqu'il s'agit de séries typiques, où les écarts ne sont pas attribuables au jeu du hasard, nous préférons, quant à nous, l'écart arithmétique moyen (d_m). Car, pour le même nombre de termes, l'un et l'autre varient avec la somme des écarts moyennaux (c'est-à-dire avec les dimensions des termes et avec leur variabilité moyennale), mais σ se trouve encore influencé par la concentration (ou l'extension) de cette somme sur un nombre plus ou moins grand de termes (c'est-à-dire par l'inégalité plus ou moins grande des écarts moyennaux). Par conséquent, tandis que le rapport $\frac{d_m}{v_m}$ nous donne effectivement l'expression de la variabilité moyennale seule, le rapport $\frac{\sigma}{v_m}$ présente encore l'effet combiné de la variabilité moyennale et de la concentration des écarts.

Pour la même raison, le rapport de l'écart moyennal quadratique moyen (σ) à l'écart moyennal arithmétique moyen (d_m) présente pour nous un intérêt tout particulier. Il est vrai que dans le rapport $\frac{\sigma}{d_m}$, le dénominateur diffère du numérateur en partie parce que le nombre des termes (n) entre dans leur calcul de façon un peu différente ; cependant, la différence essentielle entre eux provient du fait que le numérateur (σ) est influencé par la concentration des écarts moyennaux tandis que le dénominateur (d_m) ne l'est pas ; dans ce rapport, l'influence des dimensions des phénomènes envisagés, de leur variabilité moyennale (en partie aussi celle du nombre des termes de la série) se trouve éliminée tandis que celle de la concentration subsiste entière. Quand il s'agit en outre de séries homonomes (n étant invariable), l'effet de tous les facteurs autres que la concentration étant ainsi éliminé, la différence de leurs rapports respectifs $\frac{\sigma'}{d'_m}$, $\frac{\sigma''}{d''_m}$, etc. provient uniquement de la différence de concentration de l'écartement moyennal des séries comparées. Pour des séries homonomes, ce rapport que nous désignerons par

$$\sigma_d = \frac{\sigma}{d_m} \qquad [18]$$

peut donc servir *d'indice de la concentration des écarts moyennaux.* Et cela s'applique aux séries périodiques comme aux séries non-périodiques.

Nous verrons un peu plus loin (§§ 45-46) d'une façon plus précise comment cet indice varie avec la concentration des séries. Notons cependant tout de suite que ce rapport augmente avec l'augmentation de l'inégalité entre les écarts moyennaux, c'est-à-dire avec la concentration de la série.

Ainsi, si nous appliquons cet indice de concentration aux séries A et B du § 15, séries dont l'une (A) a des écarts très concentrés et dont l'autre (B) a des écarts qui ne le sont pas du tout, nous trouvons pour elles les indices de concentration des écarts (σ_d) que voici :

$$A \ldots \sigma_d = \frac{\sigma}{d_m} = \frac{\sqrt{\frac{72}{12}}}{1} = \sqrt{6} = 2{,}45$$

$$B \ldots \sigma_d = \frac{\sigma}{d_m} = \frac{\sqrt{\frac{12}{12}}}{1} = 1$$

Dans la série B, la concentration est nulle (puisque tous les écarts moyennaux sont égaux entre eux) ; et, en effet, là où tous les écarts moyennaux sont égaux, l'écart quadratique moyen (σ) est égal à l'écart arithmétique moyen (d_m) ; leur rapport devient alors $= 1$, qui marque ainsi le minimum de concentration (la concentration nulle).

Si nous appliquons l'indice de concentration σ_d aux séries A_1 et B_1 du § 16, qui sont toutes deux moins concentrées que A et plus que B et dont l'une (A_1) est plus concentrée que l'autre, nous trouvons les valeurs suivantes :

$$A_1 \ldots \sigma_d = \frac{\sigma}{d_m} = \frac{\sqrt{\frac{144}{12}}}{2} = \frac{\sqrt{12}}{2} = \sqrt{3} = 1{,}73$$

$$B_1 \ldots \sigma_d = \frac{\sigma}{d_m} = \frac{\sqrt{\frac{102}{12}}}{2} = \sqrt{\frac{17}{8}} = \sqrt{2{,}125} = 1{,}46\ .$$

§ 34. — D'ailleurs, notre indice de la concentration des écarts peut être encore simplifié. En effet,

$$\sigma_d = \frac{\sigma}{d_m} = \frac{\sqrt{\frac{\Sigma d^2}{n}}}{\frac{\Sigma d}{n}} = \sqrt{\frac{n\,\Sigma d^2}{\Sigma^2 d}}. \qquad [19]$$

Cette formule est, pour le calcul, plus commode que la formule [18] puisqu'elle remplace les deux divisions sur n par une seule multiplication par ce même nombre. Mais dans cette dernière expression de l'indice de concentration, la racine carrée est parfaitement inutile. Car $\frac{\sigma}{d_m}$ est (nous venons de le voir) égal ou supérieur à 1 ; l'expression qui se trouve sous le radical varie donc (avec la concen-

tration) dans le même sens et plus visiblement que la racine carrée de cette même expression. La racine carrée complique donc le calcul et, en même temps, rend le résultat moins expressif. Comme indice de la concentration des écarts nous pouvons donc prendre simplement l'expression qui se trouve sous le radical ; nous la désignerons par C_n. Nous pouvons dire ainsi que l'indice de la concentration des écarts des séries homonomes (C_n) s'exprime par la formule :

$$C_n = \left(\frac{\sigma}{d_m}\right)^2 = n\,\frac{\Sigma\, d^2}{\Sigma^2\, d} \qquad [20]$$

cela veut dire que *l'indice de la concentration des écarts des séries homonomes C_n est donné par la somme des carrés des écarts moyennaux divisée par le carré de la somme de ces mêmes écarts (pris en valeurs absolues) et multipliée par le nombre des termes de la série.*

Mais puisqu'il s'agit de comparer des séries homonomes, n étant constant, cet indice de concentration peut être encore simplifié par la suppression de n ; nous le désignerons par C et nous l'appelerons *indice simplifié de la concentration des écarts*, pour le distinguer du précédent (C_n) qu'on peut appeler indice amplifié de la concentration des écarts. Nous avons ainsi :

$$C = \frac{\Sigma\, d^2}{\Sigma^2\, d}, \qquad [21]$$

c'est-à-dire que *l'indice simplifié de la concentration des écarts de séries homonomes est donné par le rapport de la somme des carrés des écarts moyennaux au carré de la somme de ces écarts.*

Appliqués aux mêmes exemples concrets pris au § précédent, les indices de concentration C_n et C donnent pour les séries A et B les valeurs suivantes :

Indice amplifié de la concentration des écarts (C_n) :

$$\text{Pour } A \ldots C_n = 12 \,.\, \frac{36 + 36}{12^2} = \frac{2.36}{12} = 6$$

$$\text{Pour } B \ldots C_n = 12 \,.\, \frac{12}{12^2} = 1$$

Indice simplifié de la concentration des écarts (C) :

$$\text{Pour } A \ldots C = \frac{36 + 36}{12^2} = 0{,}50$$

$$\text{Pour } B \ldots C = \frac{12}{12^2} = \frac{1}{12} = 0{,}083\,.$$

D'après C comme d'après C_n, la série A a un indice de concentration six fois plus élevé que B.

Appliqués aux séries A_1 et B_1, les indices de concentration sont respectivement :

Indice amplifié de la concentration des écarts (C_n) :

$$\text{Pour } A_1, \ldots C_n = 12 \cdot \frac{4 \cdot 36}{24^2} = 3$$

$$\text{Pour } B_1, \ldots C_n = 12 \cdot \frac{102}{24^2} = 2.125$$

Indice simplifié de la concentration des écarts (C) :

$$\text{Pour } A_1, \ldots C = \frac{4 \cdot 36}{24^2} = \frac{1}{4} = 0{,}25$$

$$\text{Pour } B_1, \ldots C = \frac{102}{24^2} = \frac{17}{96} = 0{,}177\,.$$

Le rapport de la concentration des écarts de A_1 et de B_1 est naturellement le même, que l'on prenne C_n ou C.

§ 35. — Le rapport de l'écart *successif* quadratique moyen (ς) à l'écart successif arithmétique moyen (e_m) présente une analogie parfaite avec l'écart moyennal quadratique moyen (σ) divisé par l'écart moyennal arithmétique moyen (d_m). En effet, dans l'expression

$$\varsigma_e = \frac{\varsigma}{e_m}, \tag{22}$$

comme dans l'expression [18], le numérateur de la fraction est la moyenne quadratique des mêmes quantités dont le dénominateur

est la moyenne arithmétique. Pour des séries homonomes, la valeur de ce rapport varie uniquement avec la mesure de l'inégalité des divers écarts successifs, c'est-à-dire, pour des séries *monocéphales*, avec l'allure plus ou moins précipitée de la série (§ 17). Le rapport $\frac{\varsigma}{e_m}$ *est ainsi l'indice de la précipitation des séries homonomes monocéphales* (*).

Cet indice peut être appliqué à n'importe quelle série monocéphale, périodique ou non. Pour des séries monocéphales *périodiques*, où, comme nous l'avons vu plus haut [3], $e_m = \frac{2\alpha}{n}$, notre expression [22] prend la valeur suivante :

$$\varsigma_e = \frac{\varsigma}{2\alpha : n} = \frac{n}{2} \cdot \frac{\varsigma}{\alpha} \text{. (**)} \qquad [23]$$

Mais n étant constant (puisqu'il s'agit de comparer des séries *homonomes*), c'est donc simplement l'expression :

$$\varsigma_\alpha = \frac{\varsigma}{\alpha} \qquad [24]$$

qui peut servir d'indice de l'allure précipitée (ou, inversement, graduelle) des séries comparées.

Une fois de plus nous constatons ainsi l'intérêt des valeurs extrêmes quand il s'agit de séries de variations typiques. Et de nouveau, nous voyons l'intérêt que peut présenter, en particulier, la combinaison des deux notions fondamentales de toute variation : de la moyenne (représentée ici par ς) et des extrêmes (α). — (Cf. §§ 10, 25 et 32).

Cependant, nos indices [22], [23] et [24] peuvent, eux aussi, être simplifiés tout comme l'indice de concentration [18]. En effet,

$$\varsigma_e = \frac{\varsigma}{e_m} = \frac{\sqrt{\frac{\Sigma e^2}{n}}}{\Sigma e : n} = \sqrt{\frac{\Sigma e^2 \cdot n^2}{n \cdot \Sigma^2 e}} = \sqrt{n \frac{\Sigma e^2}{\Sigma^2 e}} \text{.} \qquad [25$$

(*) Le problème est plus complexe pour des séries polycéphales, où les écarts successifs peuvent plusieurs fois changer le sens (le signe) de leurs variations. Nous y reviendrons peut-être une autre fois.

(**) Si n est impair, il est plus simple de prendre l'expression : $\frac{n\,\varsigma}{\alpha}$.

En supprimant le radical pour les mêmes raisons que celles indiquées plus haut (§ 34), et en désignant l'indice de précipitation ainsi simplifié par P_n, nous pouvons dire :

$$P_n = n \frac{\Sigma e^2}{\Sigma^2 e} . \qquad [26]$$

Pour les séries monocéphales homonomes, n étant constant, l'indice de précipitation (P) peut encore être simplifié par la suppression de n. On obtient alors :

$$P = \frac{\Sigma e^2}{\Sigma^2 e} . \qquad [27]$$

Comme *indice de précipitation des séries monocéphales homonomes* on peut donc prendre simplement *le rapport de la somme des carrés des écarts successifs au carré de la somme de ces mêmes écarts* (pris en valeurs absolues).

Pour l'établissement de nos indices de précipitation P_n [26] et P [27], comme pour ς_e [22], nous n'avons introduit aucune considération particulière propre aux séries périodiques ; ils sont donc valables pour toutes les séries, périodiques au non. Notons cependant que dans les séries périodiques (typiques), le nombre des écarts successifs n est égal au nombre des termes de la série tandis que dans les séries non périodiques il est égal au nombre des termes *moins* 1 (puisque, dans ce dernier cas, l'écart entre le premier et le dernier terme de la série n'existe pas) (*).

A titre d'exemple, prenons d'abord deux séries monocéphales homonomes *non* périodiques ayant la même amplitude et, par suite, la même somme des écarts successifs et ne différant que par l'inégalité plus ou moins grande de ces écarts, c'est-à-dire par leur allure plus ou moins précipitée ou ondulatoire (graduelle) : l'une (X) a des hausses et des baisses bien brusques ; l'autre (Y) a au contraire une allure parfaitement graduelle, tous ses écarts successifs étant égaux entre eux.

(*) C'est pourquoi pour les séries *non* périodiques, la somme des écarts successifs est égale à $2\alpha - [v_n - v_1]$ en cas de séries monocéphales et à $2\Sigma\alpha' - [v_n - v_1]$ en cas de séries polycéphales, $[v_n - v_1]$ étant la valeur absolue de l'écart entre le dernier et le premier terme de la série.

Séries données :

$$X \ldots 10, 10, 18, 18, \ \ 6, 6,$$
$$Y \ldots 10, 14, 18, 14, 10, 6.$$

Écarts successifs :

$$X \ldots -0, +8, \quad 0, -12, \quad 0,$$
$$Y \ldots +4, +4, -4, -\ \ 4, -4.$$

Les indices de précipitation ς_e, P_n et P auront pour les deux séries les valeurs suivantes :

Série X

$$\varsigma_e = \frac{\sqrt{\frac{8^2 + 12^2}{5}}}{20 : 5} = \sqrt{2{,}60} = 1{,}61\,;$$

$$P_n = 5 \cdot \frac{8^2 + 12^2}{20^2} = 5 \cdot 0{,}52 = 2{,}60\,;$$

$$P = \frac{8^2 + 12^2}{20^2} = 0{,}52\,;$$

Série Y

$$\varsigma_e = \frac{\sqrt{\frac{5 \cdot 4^2}{5}}}{4} = 1$$

$$P_n = 5 \cdot \frac{5 \cdot 4^2}{20^2} = 1$$

$$P = \frac{5 \cdot 4^2}{20^2} = 0{,}20$$

Comme on devait s'y attendre après ce que nous avons vu plus haut pour la concentration (§§ 33 et 34), l'indice ς_e ainsi que son carré l'indice P_n deviennent $= 1$ dans le cas de la série Y qui accuse une précipitation minima (ou une gradualité parfaite). Quant à l'indice simplifié P, il est égal à l'indice amplifié P_n divisé par 5 (5 étant le nombre des écarts successifs qui dans une série non périodique est égal au nombre des termes moins 1). Le rapport des indices de la précipitation des deux séries est naturellemente le même que l'on prenne l'indice amplifié P_n ou l'indice simplifié P (en effet, $0{,}52 : 0{,}20 = 2{,}60 : 1$) ; il est différent si l'on prend l'indice ς_e.

§ 36. — Pour le cas de séries monocéphales et homonomes *périodiques*, les indices de la précipitation peuvent être calculés d'une façon plus simple encore. En effet, en substituant, d'après [2], 2α à Σe dans les formules [26] et [27], nous trouvons :

$$P_n = \frac{n \, \Sigma \, e^2}{\Sigma^2 \, e} = \frac{n \, \Sigma \, e^2}{4 \, \alpha^2} \qquad [26^{bis}]$$

et $$P = \frac{\Sigma e^2}{\Sigma^2 e} = \frac{\Sigma e^2}{4\,\alpha^2}. \qquad [27^{bis}]$$

Le chiffre 4 étant parfaitement inutile dans le dénominateur de l'indice, nous pouvons simplement le supprimer ; nous aurons ainsi comme indices de la précipitation des séries périodiques monocéphales et homonomes les expressions fort simples que voici :

$$P_{n\alpha} = \frac{n\,\Sigma e^2}{\alpha^2} \qquad [28]$$

et $$P_\alpha = \frac{\Sigma e^2}{\alpha^2}. \qquad [29]$$

Il est évident que $P_{n\alpha}$ et P_α sont respectivement 4 fois plus élevés que P_n et P ; mais le rapport des indices de séries différentes demeure naturellement le même, que l'on prenne pour indices de la précipitation P_n, P, $P_{n\alpha}$ ou P_α ; seulement l'indice P_α est le plus simple à calculer.

Ainsi, admettons que les séries X et Y du § précédent soient des séries périodiques ; leurs écarts successifs seraient donc les suivants :

$$X \ldots (+4,\ 0,\ +8,\ 0,\ -12,\ 0)$$
$$Y \ldots (+4,\ +4,\ +4,\ -4,\ -4,\ -4).$$

Les indices de précipitation seraient alors :

Série X

$$\varsigma_e = \frac{\sqrt{\dfrac{4^2+8^2+12^2}{6}}}{(4+8+12):6} = \frac{\sqrt{\dfrac{224}{6}}}{4} = \sqrt{\frac{7}{3}} = \sqrt{2{,}33} = 1{,}53\ ;$$

$$P_n = \frac{6\,(16+64+144)}{24^2} = \frac{224}{4^2 . 6} = \frac{7}{3} = 2{,}33\ ;$$

$$P = \frac{224}{24^2} = \frac{7}{18} = 0{,}39\ ;$$

$$P_{n\alpha} = \frac{6\,(16+64+144)}{(18-6)^2} = \frac{28}{3} = 9{,}33\ ;$$

$$P_\alpha = \frac{224}{(18-6)^2} = \frac{14}{9} = 1{,}56\ ;$$

Série Y

$$\varsigma_e = \frac{\sqrt{\dfrac{6 . 4^2}{6}}}{4} = \frac{4}{4} = 1\ ;$$

$$P_n = 6\,\frac{6 . 4^2}{24^2} = \frac{24^2}{24^2} = 1\ ;$$

$$P = \frac{6 . 4^2}{24^2} = \frac{1}{6} = 0{,}17\ ;$$

$$P_{n\alpha} = 6\,\frac{6 . 4^2}{(18-6)^2} = \frac{6 . 6 . 4 . 4}{12^2} = 4\ ;$$

$$P_\alpha = \frac{6 . 4^2}{(18-6)^2} = \frac{6 . 4^2}{12^2} = \frac{2}{3} = 0{,}67\ .$$

On constate ainsi que les divers indices de la précipitation de la série X (sauf ς_e) sont aux indices correspondants de la série Y comme 2,33 : 1 ; les indices basés sur ς (donc ς_e et ς_α) sont comme les racines carrées de ce rapport (1,53 : 1).

Les indices de précipitation, comme on le voit, ne sont pas tout-à-fait les mêmes pour les séries périodiques (typiques) et non-périodiques se composant des mêmes termes (ainsi, par exemple, pour la série X, P_n est respectivement $= 2,33$ au lieu de 2,60 pour le cas de série non périodique, P est $= 0,39$ au lieu de 0,52, et ainsi de suite). Cette divergence dépend de la différence entre le dernier et le premier terme de la série ; car, comme nous l'avons déjà marqué, la différence entre ces deux termes fait partie des écarts successifs, en cas de séries périodiques, mais cet écart n'existe pas pour des séries non-périodiques (*)

(*) Comme les indices de la concentration (cf. § 33), nos indices de la précipitation se trouvent sous l'influence non seulement de la précipitation même (l'inégalité des écarts successifs), mais dans une certaine mesure aussi sous celle du nombre des termes (n). C'est pourquoi nous avons bien souligné qu'il s'agit d'indices de précipitation de séries *homonomes*: n étant le même pour les séries comparées, la différence de leurs indices ne peut alors provenir que de la différence de précipitation. On commettrait donc une erreur en appliquant ces indices à des séries hétéronomes; car on attribuerait ainsi à une différence de précipitation ce qui est, au moins en partie, l'effet de la différence du nombre des termes.

Cependant, l'erreur ne serait pas de la même grandeur pour nos divers indices. Au prime abord on penserait peut-être que l'indice simplifié $P = \frac{\Sigma e^2}{\Sigma^2 e}$ $\left(\text{ou } P\alpha = \frac{\Sigma e^2}{\alpha^2}\right)$ pourrait être appliqué à des séries hétéronomes avec le moins d'erreur, sinon tout-à-fait correctement, puisque sa formule ne contient pas n ; il paraît donc insensible aux différences du nombre des termes; par contre, la formule de P_n (ou de sa racine carrée ς_e) contient n $\left(\text{étant } = \frac{n\,\Sigma e^2}{\Sigma^2 e}\right)$. Pourtant c'est le contraire qui est vrai. Il est vrai que la formule de P ne contient pas n explicitement; seulement, on peut démontrer que le rapport $\frac{\Sigma e^2}{\Sigma^2 e}$ est lui-même fonction de n, tandis que dans le rapport $\frac{\varsigma}{e_m}\left(= \frac{\sqrt{\frac{\Sigma e^2}{n}}}{\Sigma e : n}\right)$, ou dans son carré P_n, si l'action de n n'est pas éliminée, elle est pourtant, d'une façon générale, fortement atténuée. Voici d'ailleurs un exemple concret : Soit Z une série périodique monocéphale à 9 termes ayant une allure manifestement moins préci-

CHAPITRE VII.

Indices de certains caractères des séries périodiques (*suite*).

§ 37. — On peut établir un indice de *concentration* des écarts encore d'une autre façon. En effet, nous avons vu (§§ 15 et 27) que la concentration a pour corollaire un écartement moyennal du second ordre de telle nature qu'aux différences de la concentration répondent des différences de l'écart moyennal moyen du second ordre (δ_m). Naturellement, la valeur globale de cet écart dépend aussi des dimensions des phénomènes considérés et de leur variabilité moyennale. Mais en divisant cet écart (δ_m) par l'écart moyennal arithmétique moyen du premier ordre (d_m), comme nous l'avons fait pour σ_d [18], nous trouvons l'expression suivante (que nous désignerons par K) :

$$K = \frac{\delta_m}{d_m} = \frac{\Sigma\,\delta}{\Sigma\,d} \qquad [30]$$

---

pitée que la série (à 6 termes) X et plus précipitée que la série Y citée plus haut:

Z { Série donnée . . . . (5, 7, 12, 18, 20, 15, 15, 10, 9)
Écarts successifs . . (— 4, + 2, + 5, + 6, + 2, — 5, 0, — 5, — 1)

Si nous voulions comparer la précipitation de ces trois séries (qui ne sont pas homonomes) sur la base de l'indice P, nous trouverions pour la série Z

l'indice $P = \dfrac{16 + 4 + 25 + 36 + 4 + 25 + 25 + 1}{(4 + 2 + 5 + 6 + 2 + 5 + 5 + 1)^2} = \dfrac{136}{900} = 0{,}15$ (contre 0,17 pour la série Y) ; on aurait ainsi pour z une précipitation moindre que pour la série Y qui accuse une gradation parfaite, soit la précipitation minima !

Par contre, l'indice P_n donnerait pour la série Z le chiffres de $\dfrac{9 \cdot 136}{900} = 1{,}36$, contre 1 pour Y et 2,33 pour X, ce qui correspond *grosso modo* à la réalité. Il en serait de même de l'indice $\varsigma_e = \dfrac{\varsigma}{e_m}$; il donnerait $\sqrt{1{,}36} = 1{,}17$ pour Z, contre 1 pour Y et 1,53 pour X.

Une observation parfaitement analogue peut être faite aussi au sujet des indices de concentration des écarts, et naturellement tout aussi bien pour les séries non périodiques que pour les séries périodiques.

qui est indépendante des dimensions des phénomènes envisagés et de leur variabilité moyennale et qui peut, elle aussi, servir d'indice de la concentration des écarts des séries.

Notons bien qu'en cas de concentration nulle, c'est-à-dire lorsque les écarts moyennaux (du premier ordre) sont tous égaux entre eux, les écarts moyennaux du second ordre étant tous égaux à zéro, l'indice de concentration K devient lui aussi $= 0$ (le numérateur $\Sigma\,\delta$ étant $= 0$). L'indice de concentration basé sur l'écartement du second ordre (K) ne coïncide donc pas avec les indices basés sur les carrés des écarts du premier ordre établis dans le chapitre précédent et dont σ_d et C_n marquent la concentration nulle par le chiffre 1 et C par le chiffre $\frac{1}{n}$. Ainsi pour les mêmes séries A, B, A_1 et B_1 que nous avons déjà prises comme exemples plus haut (§§ 15 et 16), les indices de concentration s'expriment par les chiffres suivants :

Série	σ_d	C_n	K
A	2,45	6,—	1,67
A_1	1,73	3,—	1,34
B_1	1,46	2,13	0,84
B	1,—	1,—	0

D'après cet exemple, on voit également que l'indice K croît relativement bien plus fortement avec l'accroissement des faibles concentrations (par exemple, de B à B_1 et à A_1) et beaucoup plus faiblement pour l'accroissement des concentrations élevées (de A_1 à A) ; c'est le contraire qui est vrai pour les indices σ_d et C_n qui croissent plus faiblement avec l'accroissement des faibles concentrations (de B à B_1) et plus fortement avec celui des fortes concentrations (de A_1 à A). Et ceci n'est pas l'effet de quelque hasard : les variations des indices basés sur les *carrés* des écarts (même si l'on en tire la racine carrée) sont plus fortes pour les gros écarts que pour les petits.

Au prime abord on croirait peut-être que l'indice de concentration K, aurait une portée plus générale que les indices basés sur les carrés des écarts du premier ordre ; on croirait notamment qu'il pourrait servir également à comparer des séries hétéronomes, car le nombre des termes (n) agit sur le numérateur ($\Sigma\,\delta$) du rapport K exactement de la même façon que sur son dénominateur ($\Sigma\,d$). En réalité,

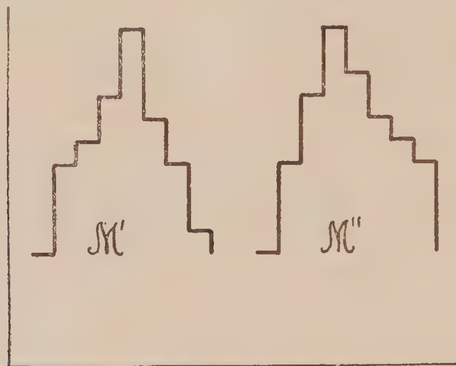
cependant, l'indice K, comme les indices de concentration dégagés plus haut, permet de comparer la concentration uniquement quand il s'agit de séries homonomes. Nous verrons plus loin (§ 59) la raison de cette restriction.

§ 38. — On sera peut-être tenté de procéder d'une façon analogue pour la *précipitation* des séries monocéphales ; on voudra peut-être prendre comme indice de la précipitation de ces séries le rapport de l'écart successif arithmétique moyen du second ordre au même écart du premier ordre $\left(\frac{\varepsilon_m}{e_m}\right)$. Car, comme nous l'avons vu (§§ 17 et 28), l'allure plus ou moins précipitée ou graduelle d'une série n'est autre chose que l'écartement successif plus ou moins inégal (ou égal) de ses termes, ce qui paraît correspondre à l'importance plus ou moins grande des écarts successifs du second ordre. Mais ceci est vrai seulement tant que le nombre des sommets (et l'importance des amplitudes) ne change pas, en particulier pour les séries monocéphales. Or, si la série des variations donnée est monocéphale, la série de ses écarts successifs ne l'est pas forcément ; elle l'est même plutôt rarement. De plus, pour *les mêmes écarts successifs* (du premier ordre), donc pour le même degré de précipitation, les écarts successifs du second ordre peuvent, en général, être bien différents selon la *disposition* des écarts successifs du premier ordre (puisque chaque écart successif du second ordre est déterminé par deux écarts voisins du premier ordre) ; pour la même précipitation, on pourrait ainsi avoir des indices de précipitation $\left(\frac{\varepsilon_m}{e_m}\right)$ différents.

Qualques exemples pour plus de clarté. Prenons deux séries périodiques (M' et M'') monocéphales, homonomes, ayant les mêmes maximum et minimum et les mêmes écarts successifs (du premier ordre), mais autrement disposés :

			Écart moyen	$\frac{\varepsilon_m}{e_m}$
M'	Série donnée (v) . . .	(5, 9, 10, 12, 15, 11, 9, 6)	—	—
	Écarts successifs (pris en valeur absolue) :			
	du 1er ordre (e) . .	(1, 4, 1, 2, 3, 4, 2, 3)	2,50	0,70
	du 2d ordre (ε) . .	(2, 3, 3, 1, 1, 1, 2, 1)	1,75	

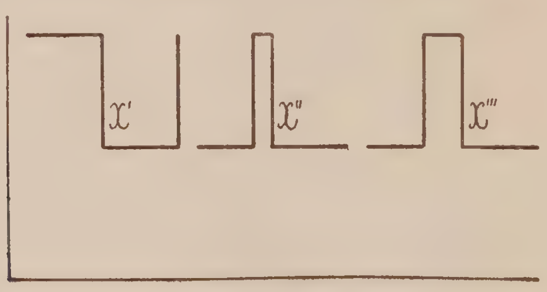
			Écart moyen	$\frac{\varepsilon_m}{e_m}$
	Série donnée (v) . . .	(5, 9, 12, 15, 13, 11, 10, 9)	—	—
	Écarts successifs (pris			
M''	en valeur absolue) :			
	du 1er ordre (e) . .	(4, 4, 3, 3, 2, 2, 1, 1)	2,50	0,30
	du 2d ordre (ε) . .	(3, 0, 1, 0, 1, 0, 1, 0)	0,75	



Graphique VII.

La série M', dont l'allure est tout aussi graduelle ou précipitée que celle de la série M'' puisqu'elle accuse exactement les mêmes écarts successifs, aurait ainsi un indice de précipitation plus que double uniquement par suite du fait que ces écarts sont dans M' autrement disposés que dans M''.

Un autre exemple : prenons trois séries périodiques monocéphales, homonomes (X', X'' et X''') qui ont la même amplitude et qui accusent les mêmes écarts successifs allant d'un coup du maximum de la série à son minimum et vice versa, mais qui diffèrent par la situation respective de la hausse et de la baisse :



Graphique VIII.

			Écart moyen	$\frac{\varepsilon_m}{e_m}$
X'	Série donnée (v) . . .	(13, 13, 13, 13, 7, 7, 7, 7)	—	—
	Écarts successifs (pris en valeur absolue) :			
	du 1er ordre (e) . .	(6, 0, 0, 0, 6, 0, 0, 0)	1,5	2
	d 2d ordre (ε) . . .	(6, 6, 0, 0, 6, 6, 0, 0)	3,0	
X''	Série donnée (v) . . .	(7, 7, 7, 13, 7, 7, 7, 7)	—	—
	Écarts successifs (pris en valeur absolue) :			
	du 1er ordre (e) . .	(0, 0, 0, 6, 6, 0, 0, 0)	1,5	1
	du 2d ordre (ε) . .	(0, 0, 0, 6, 0, 6, 0, 0)	1,5	
X'''	Série donnée (v) . .	(7, 7, 7, 13, 13, 7, 7, 7)	—	—
	Écarts successifs (pris en valeur absolue) :			
	du 1er ordre (e) .	(0, 0, 0, 6, 0, 6, 0, 0)	1,5	2
	du 2d ordre (ε) . .	(0, 0, 0, 6, 6, 6, 6, 0)	3,0	

Malgré le fait que dans les trois séries nous trouvons une seule hausse et une seule baisse de dimensions identiques pour toutes les trois, l'indice de précipitation basé sur l'écartement successif du second ordre serait pour la série X'' deux fois plus petit que pour X' et X''', et cela uniquement parce que dans X'' la hausse et la baisse se suivent immédiatement. (On remarquera aussi que pour X'', la série des écarts successifs du 1er ordre demeure monocéphale ; pour X' et X''', elle est bicéphale).

Le rapport de l'écart successif arithmétique moyen du second ordre au même écart du premier ordre (analogue à celui établi plus haut, sur la base de l'écartement moyennal, pour la concentration) ne peut donc pas servir d'indice de la précipitation des séries.

§ 39. — Nous avons vu plus haut (§ 30) que ce qui détermine la déviation de la symétrie, la *dissymétrie* des séries, et qui varie avec cette dissymétrie, ce sont les différences des écarts des termes équidistants (*), en cas de symétrie positive, et les sommes algébriques de

---

(*) Par l'expression « termes equidistants » nous entendrons toujours les termes également distants du centre de symétrie.

ces écarts en cas de symétrie négative ; nous avons vu encore que ces écarts peuvent être moyennaux, centraux ou successifs. Mais la valeur brute des écarts des termes équidistants varie non seulement avec le degré d'asymétrie de la série, mais encore (comme celle des écarts en général) avec la grandeur des termes de la série (dimensions des phénomènes considérés) et avec leur variabilité ; le total des différences (ou des sommes algébriques) des écarts de tous les couples de termes équidistants varie naturellement encore avec le nombre des termes de la série. Pour éliminer l'action de tous les facteurs autres que la dissymétrie, nous diviserons la somme des valeurs absolues des différences (ou des sommes algébriques) des écarts des termes équidistants par la somme des valeurs absolues des écarts de tous les termes de la série (§ 31). Naturellement, s'il s'agit de la dissymétrie des écarts moyennaux, il faudra diviser la somme des valeurs absolues des différences (ou des sommes algébriques) des écarts moyennaux des termes équidistants ($\Sigma\,[d' \mp d'']$) (*) par la somme des valeurs absolues des écarts moyennaux de tous les termes ($\Sigma\,[d]$) ; s'il s'agit de la symétrie des écarts successifs, il faudra diviser la somme des valeurs absolues des différences (ou des sommes algébriques) des écarts successifs des termes équidistants ($\Sigma\,[e' \mp e'']$) par la somme des valeurs absolues de tous les écarts successifs de la série ($\Sigma\,[e]$) ; s'il s'agit de la symétrie des écarts centraux, la somme des valeurs absolues des différences (ou des sommes algébriques) des écarts centraux des termes équidistants ($\Sigma\,[\eta' \mp \eta'']$) sera divisée par la somme des valeurs absolues des écarts centraux de tous les termes de la série ($(\Sigma_\eta)\,[\eta]$).

Les *indices de dissymétrie* (que nous désignerons par D) seront donc les suivants (**) :

	Dissymétrie positive	Dissymétrie négative	
Dissymétrie des écarts moyennaux :	$D_d^+ = \dfrac{\Sigma\,[d' - d'']}{\Sigma\,[d]}$;	$D_d^- = \dfrac{\Sigma\,[d' + d'']}{\Sigma\,[d]}$	[31]
Dissymétrie des écarts successifs :	$D_e^+ = \dfrac{\Sigma\,[e' - e'']}{\Sigma\,[e]}$;	$D_e^- = \dfrac{\Sigma\,[e' + e'']}{\Sigma\,[e]}$	[32]
Dissymétrie des écarts centraux :	$D_\eta^+ = \dfrac{\Sigma\,[\eta' - \eta'']}{\Sigma\,[\eta]}$;	$D_\eta^- = \dfrac{\Sigma\,[\eta' + \eta'']}{\Sigma\,[\eta]}$	[33]

(*) Dans le reste de ce chapitre, les crochets [ ] serviront à désigner la valeur absolue des expressions algébriques.

(**) Convenons d'appeler dissymétrie positive la déviation de la symétrié positive et dissymétrie négative la déviation de la symétrie négative.

Et en général, en désignant les écarts de quelque nature qu'ils soient (moyennaux, successifs ou centraux) par E, les écarts des termes situés d'un côté du centre de symétrie par E' et de l'autre par E'', nous pouvons écrire que l'indice de dissymétrie (D) est égal à :

$$D = \frac{\Sigma [E' \mp E'']}{\Sigma [E]}, \qquad [34]$$

étant bien entendu qu'il faut prendre $[E' - E'']$ pour déterminer la déviation de la symétrie positive et $[E' + E'']$ pour la déviation de la symétrie négative.

§ 40. — Dans le cas de symétrie positive parfaite, c'est-à-dire si les écarts des termes équidistants sont égaux, toutes les différences $E' - E''$ sont $= 0$ et notre indice de dissymétrie devient à son tour $= 0$. Il en est de même en cas de symétrie négative parfaite : chaque E' étant $= - E''$, toutes les sommes $E' + E''$ deviennent $= 0$ de même que l'indice de dissymétrie.

Dans le cas de symétrie partielle, notre indice de dissymétrie D est une fraction (positive) inférieure à 1. En cas de déviation de la symétrie positive et tant que les E' et les E'' conservent les mêmes signes, il est évident que pour tout couple de termes équidistants $[E' - E''] < ([E'] + [E''])$, c'est-à-dire $\frac{\Sigma [E' - E'']}{\Sigma [E'] + \Sigma [E'']} < 1$; cette expression augmente à mesure que E'' se rapproche de 0 ; et lorsque tous les E'' sont 0, l'indice D^+ devient $= 1$. Il en est de même lorsque tous les E'' ont des signes opposés à ceux de E' ; l'indice de la déviation de la symétrie positive devient alors aussi $= 1$. (En effet, dans chaque couple de termes équidistants $([E' - (- E'')] = [E'] + [E'']$; d'où : $D^+ = \frac{\Sigma [E' - E'']}{\Sigma [E]} = \frac{\Sigma [E'] + \Sigma [E'']}{\Sigma [E'] + \Sigma [E'']} = 1$). En d'autres termes : *l'indice de dissymétrie positive est une fraction variant entre* 0 *et* 1 *; il devient* 0 *lorsque la symétrie positive est parfaite* (lorsque la dissymétrie est nulle) *; il devient* 1 *lorsque toute trace de symétrie positive disparaît soit que dans tous les couples des termes équidistants l'un des termes demeure invariable* (écart $= 0$) *tandis que l'autre varie, soit encore que l'un varie en sens opposé aux variations de l'autre.*

Par un raisonnement analogue, on *arrive au même résultat pour l'indice de la dissymétrie négative* : lui aussi varie entre 0 et 1 ; à l'absence de toute déviation de la parfaite symétrie négative, correspond l'indice 0 ; l'indice augmente avec l'augmentation de cette déviation, et quand les termes équidistants n'accusent aucune variation en sens opposés (si l'un des termes de chaque couple varie tandis que l'autre reste invariable ou varie dans le même sens), l'indice de la déviation de la symétrie négative, l'indice de la dissymétrie négative, devient = 1.

Notons que dans l'établissement de notre indice de dissymétrie D nous n'avons introduit aucune condition particulière propre aux séries périodiques ; il est donc applicable à n'importe quelles séries, périodiques ou non. Dans les calculs concrets, il ne faut pourtant pas perdre de vue que pour les séries non-périodiques le nombre des écarts successifs et centraux est de 1 inférieur au nombre des termes de la série (donc $= n - 1$) tandis que pour les séries périodiques (typiques), le nombre de toutes sortes d'écarts (moyennaux, successifs ou centraux) est égal à celui des termes de la série ($= n$).

§ 41. — Pour l'application pratique de notre indice de dissymétrie on peut recommander le procédé suivant : on regarde les signes des écarts des termes équidistants ; s'ils sont systématiquement (ou à peu près) les mêmes, on applique l'indice de déviation de la symétrie positive (D^+) ; s'ils sont systématiquement opposés on appliquera l'indice de dissymétrie négative (D^-) ; si le cas paraît de prime abord douteux, on calculera les deux indices et l'on retiendra de préférence le moindre, c'est-à-dire celui qui marque la moindre déviation de la symétrie.

Appliquons maintenant effectivement ces indices à quelques exemples concrets. Prenons d'abord les deux séries non périodiques du § 18 et calculons les indices de dissymétrie de leurs écarts centraux.

							Terme central
Series données (v)	I . .	20,	5,	10,	5,	20	10
	II . .	4,	3,	5,	7,	6	5
Écarts centraux (η)	I . .	+ 10,	— 5,	0,	— 5,	+ 10	
	II . .	— 1,	— 2,	0,	+ 2	+ 1	

Pour la série I (les signes des écarts des termes équidistants étant partout les mêmes), il s'agit d'un indice de déviation de la symétrie positive ; il est égal à :

$$D^{+}_{\eta} = \frac{\Sigma\,[\eta' - \eta'']}{\Sigma\,[\eta]} = \frac{[10 - 10] + [-5 - (-5)]}{10 + 5 + 5 + 10} = \frac{0}{30} = \mathbf{0}\,,$$

ce qui correspond à une dissymétrie nulle, soit à une parfaite symétrie positive des écarts centraux.

Pour la série II (les signes des termes équidistants étant partout opposés), il s'agit de fixer l'indice de déviation de la symétrie négative, il est égal à :

$$D^{-}_{\eta} = \frac{\Sigma\,[\eta' + \eta'']}{\Sigma\,[\eta]} = \frac{[-1 + 1] + [-2 + 2]}{1 + 2 + 2 + 1} = \frac{0}{28} = 0\,,$$

ce qui marque une dissymétrie négative nulle, soit une parfaite symétrie négative des écarts centraux.

On pourrait de même calculer pour les mêmes séries les indices de dissymétrie des écarts moyennaux ou successifs.

Calculons maintenant les indices de dissymétrie des écarts *moyennaux* des quatre séries ayant les mêmes termes, mais autrement disposés, mentionnées à titre d'exemple au § 29.

Séries données (v) :

								Moyenne arithmétique ($\overline{v_m}$)
I . . .	8,	5,	3,	10,	17,	15,	12	10
II . . .	12,	3,	15,	10,	17,	5,	8	10
III . . .	17,	8,	15,	10,	3,	5,	12	10
IV . . .	17,	3,	10,	15,	12,	8,	5	10

Écarts moyennaux (d)

I . . .	— 2,	— 5,	— 7,	0,	+ 7,	+ 5,	+ 2
II . . .	+ 2,	— 7,	+ 5,	0,	+ 7,	— 5,	— 2
III . . .	+ 7,	— 2,	+ 5	0,	— 7,	— 5,	+ 2
IV . . .	+ 7,	— 7,	0,	+ 5,	+ 2,	— 2,	— 5

Série I. — Les écarts des termes équidistants ont toujours des signes opposés. L'indice de dissymétrie négative est :

$$D_d^- = \frac{[-2+2]+[-5+5]+[-7+7]}{2+5+7+7+5+2} = \frac{0}{28} = 0 .$$

La série I présente donc une parfaite symétrie négative des écarts moyennaux.

Série II. — Les écarts des termes équidistants ont plutôt les mêmes signes. L'indice de déviation de la symétrie positive est :

$$D_d^+ = \frac{[2-(-2)]+[-7-(-5)]+[5-7]}{2+5+7+7+5+2} = \frac{4+2+2}{28} = \frac{8}{28} = 0{,}29$$

Cette série présente ainsi une sensible déviation de la symétrie positive des écarts moyennaux.

Série III. — Les écarts des termes équidistants se font tantôt dans le même sens tantôt dans un sens opposé. Calculons les indices de déviation de la symétrie aussi bien positive que négative.

Indice de dissymétrie positive :

$$D_d^+ = \frac{[7-2]+[-2+5]+[5+7]}{28} = \frac{5+3+12}{28} = \frac{20}{28} = 0{,}71$$

Indice de dissymétrie négative :

$$D_d^- = \frac{[7+2]+[-2-5]+[5-7]}{28} = \frac{9+7+2}{28} = \frac{18}{28} = 0{,}64 .$$

La série est très asymétrique, mais la déviation est moins forte pour la symétrie négative.

Série IV. — Les signes des écarts des termes équidistants sont plutôt opposés. L'indice de la dissymétrie négative est :

$$D_d^- = \frac{[7-5]+[-7-2]+[0-2]+[5]}{7+7+5+2+2+5} \text{ (*) } =$$

$$= \frac{2+9+2+5}{28} = \frac{18}{28} = 0{,}64 .$$

(*) Dans le cas de symétrie partielle des écarts moyennaux (c'est alors seulement que le problème peut se poser), lorsque le terme central s'écarte lui-même de la moyenne, il paraît juste de compter cet écart parmi les manifestations de

Prenons encore le cas de la série citée à titre d'exemple au § 22, 1°, et calculons l'indice de dissymétrie de ses écarts *successifs* (en partant du terme central 15).

Série donnée (v) . .	3,	5,	16,	20,	15,	6,	4,	2,	1
Écarts successifs (e)	— 2,	— 11,	— 4,	+ 5,	—,	— 9,	— 2,	— 2,	— 1

On voit d'emblée que la série est très asymétrique par rapport au terme median 15, pris comme centre de symétrie. En effet, l'indice de dissymétrie positive est égal à :

$$D_e^+ = \frac{[5+9]+[-4+2]+[-11+2]+[-2+1]}{2+11+4+5+9+2+2+1} =$$

$$= \frac{14+2+9+1}{36} = \frac{26}{36} = 0{,}72 \, .$$

L'indice de dissymétrie négative est tout aussi élevé :

$$D_e^- = \frac{[5-9]+[-4-2]+[-11-2]+[-2-1]}{2+11+4+5+9+2+2+1} =$$

$$= \frac{4+6+13+3}{36} = \frac{26}{36} = 0{,}72 \, .$$

Par contre, si nous prenons comme centre de symétrie non pas le terme médian 15, mais le terme 20, les écarts successifs en partant de ce centre sont :

$$-2, \; -11, \; -4, \; -, \; -5, \; -9, \; -2, \; -2, \; -1.$$

Les termes équidistants des deux côtés du terme 20 pris comme centre de symétrie accusent des écarts ayant tous les mêmes signes ;

---

dissymétrie (donc aussi bien au numérateur qu'au dénominateur de l'indice de dissymétrie). Nous attirons pourtant l'attention sur le fait que plus haut (§ 19, série I) nous avons pris comme exemple de parfaite symétrie positive des écarts moyennaux une série dont le terme central s'écartait de la moyenne. Il y a là un problème qu'il faudrait encore élucider.

ces écarts accusent ainsi une certaine symétrie positive ; l'indice de déviation de la symétrie positive est pour eux égal à :

$$D_e^+ = \frac{[-4+5]+[-11+9]+[-2+2]+[-2]+[-1]}{2+11+4+5+9+2+2+1} (*) =$$

$$= \frac{1+2+2+1}{36} = \frac{6}{36} = 0,17 \,.$$

La série, qui est très asymétrique par rapport à son terme médian, se montre donc, au contraire, fort peu asymétrique ou, si l'on préfère, fort symétrique par rapport au terme 20 pris comme centre de symétrie.

Si la série indiquée était une série *périodique*, on pourrait, en outre, la commencer par le terme 1. Si l'on prend alors le terme 20 comme centre de symétrie, la série même et ses écarts successifs prendront la forme suivante :

Série donnée (v) . . (1, 3, 5, 16, 20, 15, 6, 4, 2)

Écarts successifs (e) (— 2, — 2, — 11, — 4, —, — 5, — 9, — 2, — 2, — 1).

L'indice de dissymétrie (positive) des écarts successifs est ici :

$$D_e^+ = \frac{[-4+5]+[-11+9]+[-2+2]+[-2+2]+[-1]}{2+2+11+4+5+9+2+2+1} =$$

$$= \frac{1+2+1}{38} = \frac{4}{38} = 0,11 \,.$$

La série périodique considérée présenterait ainsi une symétrie positive des écarts successifs extrêmement élevée.

Comme on le voit d'après ces divers exemples, le calcul de notre indice de dissymétrie D constitue une opération des plus simples.

(*) Les écarts successifs qui n'ont pas leur pendant de l'autre côté du centre de symétrie sont entièrement des manifestations de dissymétrie et, par suite simplement additionnés (en valeur absolue) aux différences des écarts des termes équidistants dont la somme forme le numérateur de l'indice de dissymétrie. (Cf. note précédente). On procèdera de la même façon en cas de dissymétrie négative de ces écarts. Ce procédé n'est cependant pas exempt de toute critique, et le résultat auquel il aboutit doit être considéré plutôt comme approximatif.

§ 42. — Dans notre indice de dissymétrie D, on pourrait remplacer les valeurs absolues par les carrés des mêmes quantités ; on arriverait ainsi à une expression qu'on pourrait appeler *l'indice carré* de la dissymétrie (qu'il ne faut pas confondre avec le carré de l'indice), que nous désignerons par ${}^2D$ et qui est égal à :

$$ {}^2D = \frac{\Sigma\,(E' \mp E'')^2}{\Sigma\, E^2}\,, \qquad [35]$$

étant toujours entendu qu'il faut prendre $(E' - E'')^2$ pour déterminer la dissymétrie positive et $(E' + E'')^2$ pour la dissymétrie négative.

Pour atténuer l'action de l'élévation au carré, on pourrait prendre la racine carrée de l'indice carré ; on obtiendrait ainsi un indice qu'on pourrait appeler *quadratique*, que l'on désignerait par ${}^{\sqrt{2}}D$ et qui s'exprimerait par la formule suivante :

$$ {}^{\sqrt{2}}D = \sqrt{\frac{\Sigma\,(E' - E'')^2}{\Sigma\, E^2}}\,. \qquad [36]$$

Lorsque la symétrie est parfaite, l'indice carré de la dissymétrie (${}^2D$) est égal à 0. En effet, en cas de symétrie positive parfaite, tous les E'' sont égaux aux E' correspondants et $\Sigma\,(E' - E'')^2$ ainsi que l'indice ${}^2D+$ deviennent $= 0$. Il en est de même pour l'indice carré de dissymétrie négative en cas de dissymétrie négative parfaite, lorsque tous les $E'' = - E'$: leur somme $E' + E''$ et l'indice sont aussi $= 0$. — Dans les cas où dans tous les couples de termes équidistants, l'un des termes présente quelque écart tandis que l'autre n'en présente aucun, c'est-à-dire si E' a quelque valeur positive ou négative tandis que E'' est égal à 0 (ou vice versa), l'indice carré de la dissymétrie ${}^2D = \frac{\Sigma\,(E')^2}{\Sigma\,(E')^2} = 1$. — Dans le cas où tous les E'' seraient $= - E'$, 'indice carré de dissymétrie positive ${}^2D+ = \frac{\Sigma\,(E' - E'')^2}{\Sigma\,(E') + \Sigma\,(E')^2} = \frac{\Sigma\,(2\,E')^2}{2\,\Sigma\,(E')^2} =$ $= \frac{4\,\Sigma\,(E')^2}{2\,\Sigma\,(E')^2} = 2$. De même dans le cas où tous les $E'' = E'$, l'indice carré de dissymétrie négative serait : ${}^2D- = \frac{\Sigma\,(E' + E'')^2}{\Sigma\, E^2} = \frac{4\,\Sigma\,(E')^2}{2\,\Sigma\,(E')^2} = 2$. — Quant à l'indice quadratique de la dissymétrie, il serait égal à a racine carrée des expressions obtenues pour l'indice carré correspondant.

En résumé, les indices de la dissymétrie, l'indice simple (D), carré (^2D) et quadratique $\left(\sqrt[\sqrt{2}]{D}\right)$, prendraient les valeurs suivantes :

	Dissymétrie positive			Dissymétrie négative		
	$D+$	$^2D+$	$\sqrt[\sqrt{2}]{D}+$	$D-$	$^2D-$	$\sqrt[\sqrt{2}]{D}-$
	—	—	—	—	—	—
En cas où tous les :						
$E'' = E'$	0	0	0	1	2	$\sqrt{2}$
E'' (ou E') $= 0$ . .	1	1	1	1	1	1
$E'' = -E'$	1	2	$\sqrt{2}$	0	0	0

La principale différence entre l'indice simple (D) basé sur les valeurs absolues des différences (ou des sommes algébriques) des écarts et les indices $\left({}^2D \text{ et } \sqrt[\sqrt{2}]{D}\right)$ basés sur leurs carrés peut être constatée dans les cas où E' et E'' suivent des sens différents de celui de la symétrie supposée, c'est-à-dire en cas de $E'' = -E'$ pour la dissymétrie positive et de $E'' = E'$ pour la dissymétrie négative. Dans le cas extrême, où tous les $E'' = -E'$, $^2D+$ monte à 2 et $\sqrt[\sqrt{2}]{D}+$ monte à $\sqrt{2}$ ($= 1{,}41$) tandis que $D+$ ne s'élève jamais au-dessus de 1 ; il en est de même pour $D-$, $^2D-$, et $\sqrt[\sqrt{2}]{D}-$ si $E' = E''$. Ainsi, tandis que D varie entre 0 et 1, ^2D varie entre 0 et 2 et $\sqrt[\sqrt{2}]{D}$ varie entre 0 et $\sqrt{2}$.

Quant à nous, et jusqu'à nouvel avis, nous ne voyons pas quel avantage il y aurait de compliquer sensiblement le calcul pour établir un indice carré (^2D) ou quadratique $\left(\sqrt[\sqrt{2}]{D}\right)$ à la place du simple indice de dissymétrie D. Nous avons cependant tenu à signaler aussi les deux autres indices possibles, ne serait-ce que sous bénéfice d'inventaire.

CHAPITRE VIII.

Coefficients de certains caractères des séries périodiques.

§ 43. — Notre indice de dissymétrie D nous paraît présenter des avantages incontestables sur les indices proposés jusqu'ici, notamment sur la moyenne cubique des écarts moyennaux. D'abord et surtout, notre indice nous semble beaucoup plus correct. Il est, de plus, simple à calculer. Puis il permet de distinguer non seulement la mesure de la dissymétrie, mais aussi les divers genres de dissymétrie (et de symétrie). Enfin, de la mesure de la dissymétrie il donne une notion beaucoup plus nette. En effet, les indices de dissymétrie proposés jusqu'ici, et en particulier l'écart moyennal cubique moyen, n'ont qu'une seule limite visible, la limite inférieure, qui est $= 0$ pour la dissymétrie nulle ; ils n'ont aucune limite visible supérieure. Il s'ensuit que (même abstraction faite des graves défauts que nous avons indiqués au § 29) ces indices permettent tout au plus de dire dans quelle série la dissymétrie est *plus forte* et dans quelle autre elle est *plus faible* ; mais pour une série prise séparément, l'indice de dissymétrie ne nous dit pas encore si cette dissymétrie est *forte* ou *faible*. Ainsi, de deux indices qu'on aurait obtenus par ces méthodes et dont l'un serait de 0,9 et l'autre de 1,2, nous pourrions tout au plus conclure que le second marque une plus forte dissymétrie ; mais ce second lui-même, égal à 1,2, marque-t-il une dissymétrie excessivement élevée, notable ou à peu près négligeable ? — nous n'en saurions encore rien. Par contre, notre indice D, ne pouvant varier qu'entre les deux limites de 0 au minimum et de 1 au maximum, marque par sa valeur numérique (0,90, 0,60, 0,10, etc.) le *degré* de la dissymétrie d'une série même prise isolément ; il marque ce degré eu égard aux limites possibles de la dissymétrie. Il se rapproche ainsi d'un vrai *coefficient* de la dissymétrie.

Par contre, ce grand avantage n'existe pas pour nos autres indices dégagés plus haut, en particulier pour les indices de la concentration (pour l'indice K comme pour les indices du type C) ainsi que pour celui de la précipitation (P). Pour les indices C_n et P_n, nous avons la limite inférieure qui est égale à 1 (et à $\frac{1}{n}$ pour les indices simplifiés C et P) ; pour l'indice K, la limite inférieure est $= 0$; mais nous

ignorons encore la limite supérieure de chacun de ces indices. Dans ces conditions, ces indices demeurent *inachevés* : ils permettent tout au plus de dire laquelle des séries est plus concentrée et laquelle l'est moins, laquelle a une allure plus précipitée et laquelle plus ondulatoire, mais nous ne savons pas si tel ou tel chiffre signifie un degré élevé ou bas de précipitation ou de concentration. Nous devons donc chercher si nos indices ont une limite supérieure et, s'ils en ont une, la fixer. Nous devons, en général, préciser davantage la mesure des variations de nos indices en fonction des variations de l'intensité des caractères auxquels ils correspondent.

§ 44. — Mais avant de préciser la nature et les limites des variations de nos indices et pour pouvoir le faire, nous nous permettons de rappeler ici trois théorèmes relatifs aux maxima et aux minima.

1° *La somme des carrés de plusieurs variables* (v), *ayant les mêmes signes et dont la somme* (S) *est constante, devient maxima quand tous les variables sauf un deviennent* $= 0$ (le seul variable valable devenant alors $= S$) ; *elle est alors égale au carré de la somme de ces variables.* — Nous désignerons cette proposition sous la forme suivante :

$$\sum_{\max} v^2 = S^2 \tag{37}$$

2° *La somme des carrés de* n *variables, dont la somme* (S) *et le nombre* (n) *sont constants, devient minima quand tous les variables deviennent égaux entre eux* $\left(\text{soit égaux chacun à } \frac{S}{n}\right)$; *elle est alors égale au carré de la somme des variables divisé par leur nombre,* soit :

$$\sum_{\min} v^2 = \frac{S^2}{n} \tag{38}$$

3° *Le produit* (P) *de deux facteurs variables dont la somme* (S) *est constante, devient maximum quand ces deux facteurs sont égaux* $\left(\text{c'est-à-dire quand chacun d'eux est } = \frac{S}{2}\right)$; *il est alors égal au carré de la demi-somme de ces variables,* soit :

$$P_{\max} = \left(\frac{S}{2}\right)^2 \tag{39}$$

Nous nous permettons également de donner ici une démonstration de ces propositions par des procédés élémentaires, la démonstration même pouvant être utile à nos déductions ultérieures.

Proposition 1. — On sait que le carré de la somme de plusieurs quantités est égal à la somme des carrés de ces quantités plus le double de la somme des produits de ces quantités prises deux à deux [$(a + b)^2 = a^2 + b^2 + 2\,a\,b$; $(a + b + c)^2 = a^2 + b^2 + c^2 + 2\,(ab + ac + bc)$, et ainsi de suite]. Nous pouvons ainsi écrire : $(\Sigma\, v)^2 = \Sigma\, v^2 + 2\, \Sigma\, v\, v$ (par $v\, v$ nous désignons les produits de ces quantités prises deux à deux) ; d'où il résulte :

$$\Sigma\, v^2 = (\Sigma\, v)^2 - 2\,\Sigma\, v\, v = S^2 - 2\,\Sigma\, v\, v \qquad [40]$$

Les quantités envisagées étant du même signe, leurs produits deux à deux sont tous positifs ; $2\,\Sigma\, v\, v$ est donc une quantité positive. S étant constant, l'expression $(S^2 - 2\,\Sigma\, v\, v)$ est donc *maxima* quand $\Sigma\, v\, v = 0$, ce qui n'est possible que lorsque tous les $v\, v$ sont $= 0$. Si la somme de ces quantités (S) n'est pas égale à 0, leurs produits ne peuvent devenir $= 0$ que si toutes ces quantités sauf une sont $= 0$; la quantité restant seule valable est alors $= S$ et $(\Sigma\, v)^2 = S^2$.

Proposition 2. — Prenons plusieurs quantités $v_1, v_2, v_3, \ldots v_n$; désignons leur moyenne arithmétique par v_m et leurs écarts à cette moyenne par $d_1, d_2, d_3, \ldots d_n$ (pouvant être positifs ou négatifs) ; nous pouvons mettre ainsi : $v_1 = v_m + d_1$; $v_2 = v_m + d_2$; $v_3 = v_m + d_3$; $\ldots v_n = v_m + d_n$.

Si $n = 2$, la somme des carrés des quantités envisagées sera :

$$v_1^2 + v_2^2 = (v_m + d_1)^2 + (v_m + d_2)^2 = v_m^2 + 2\,v_m\,d_1 + d_1^2 + v_m^2 + 2\,v_m\,d_2 + d_2^2 = 2\,v_m^2 + 2\,v_m\,(d_1 + d_2) + (d_1^2 + d_2^2)\,.$$

Si $n = 3$, la somme des carrés des quantités envisagées sera :

$$\begin{aligned}(v_1^2 + v_2^2) + v_3^2 = &\left\{\begin{array}{l} 2\,v_m^2 + 2\,v_m\,(d_1 + d_2) + (d_1^2 + d_2^2) \\ + v_m^2 + 2\,v_m\,d_3 \qquad\qquad + d_3^2 \end{array}\right. \\ &= 3\,v_m^2 + 2\,v_m\,(d_1 + d_2 + d_3) + (d_1^2 + d_2^2 + d_3^2)\end{aligned}$$

Si $n = 4$, la somme des carrés des quantités envisagées sera :

$$\begin{aligned}(v_1^2 + v_2^2 + v_3^2) + v_4^2 = &\left\{\begin{array}{l} 3\,v_m^2 + 2\,v_m\,(d_1 + d_2 + d_3) + (d_1^2 + d_2^2 + d_3^2) \\ + v_m^2 + 2\,v_m\,d_4 \qquad\qquad\qquad + d_4^2 \end{array}\right. \\ &= 4\,v_m^2 + 2\,v_m\,(d_1 + d_2 + d_3 + d_4) + (d_1^2 + d_2^2 + d_3^2 + d_4^2)\end{aligned}$$

Et ainsi de suite. En général quel que soit le nombre (n) des quantités envisagées, la somme de leurs carrés sera ainsi :

$$v_1^2 + v_2^2 + v_3^2 + \ldots + v_n^2 = n\, v_m^2 + 2\, v_m\, (d_1 + d_2 + d_3 + \ldots + d_n) + \\ + (d_1^2 + d_2^2 + d_3^2 + \ldots + d_n^2)\,.$$

Mais la somme algébrique de tous les écarts à la moyenne arithmétique $(d_1 + d_2 + d_3 + \ldots + d_n)$ est toujours $= 0$; nous pouvons donc écrire :

$$v_1^2 + v_2^2 + v_3^2 + \ldots + v_n^2 = n\, v_m^2 + (d_1^2 + d_2^2 + d_3^2 + \ldots + d_n^2)\,,$$

soit d'après notre notation habituelle :

$$\Sigma\, v^2 = n\, v_m^2 + \Sigma\, d^2\,. \qquad [41]$$

Mais $v_m = \dfrac{S}{n}$; nous pouvons donc écrire :

$$\Sigma\, v^2 = n \left(\frac{S}{n}\right)^2 + \Sigma\, d^2 = \frac{S^2}{n} + \Sigma\, d^2 \qquad [42]$$

S et n étant constants et $\Sigma\, d^2$ étant toujours positif, $\Sigma\, v^2$ est *minimum* quand $\Sigma\, d^2 = 0$. Mais la somme des quantités positives $(\Sigma\, d^2)$ ne peut être $= 0$ que si toutes ces quantités sont égales à 0, c'est-à-dire lorsque les d dont tous $= 0$; tous les v sont alors égaux à v_m et égaux entre eux et $\Sigma\, v^2 = \dfrac{S^2}{n} + 0 = \dfrac{S^2}{n}$.

Proposition 3. — Soit v_1 et v_2, deux quantités dont la somme est égale à une constante S ; soit $v_1 = \dfrac{S}{2} + d$ et $v_2 = \dfrac{S}{2} - d$ (d étant une certaine quantité positive ou négative). Le produit (P) de ces deux quantités sera :

$$P = v_1\, v_2 = \left(\frac{S}{2} + d\right)\left(\frac{S}{2} - d\right) = \left(\frac{S}{2}\right)^2 - d^2\,. \qquad [43]$$

$\dfrac{S}{2}$ étant constant et d^2 étant toujours positif, le produit P est maxi-

mum quand $d^2 = 0$, c'est-à-dire quand $v_1 = v_2 = \frac{S}{2}$; leur produit est alors égal à $\left(\frac{S}{2}\right)^2$.

Ce qu'il fallait démontrer (*).

§ 45. — Ainsi donc la somme des carrés de n variables ayant les mêmes signes et dont la somme est constante et égale à S, atteint son minimum quand tous ces variables sont égaux entre eux (proposition 2) elle augmente avec l'augmentation de l'inégalité entre eux, plus exactement avec l'augmentation des écarts (d) entre eux et leur moyenne arithmétique [42], et elle atteint son maximum quand un des variables est à lui seul égal à la somme S, tous les autres variables devenant $= 0$ (proposition 1). A son minimum, la somme des carrés des n variables est $= \frac{S^2}{n}$ [d'après 38] ; à son maximum elle est $= S^2$ [d'après 37]. Nous pouvons ainsi écrire :

$$S^2 \overline{>} \Sigma v^2 \overline{>} \frac{S^2}{n} \tag{44}$$

Prenons maintenant un cas plus particulier : prenons le cas où les variables envisagées sont les valeurs absolues des écarts moyennaux (d) d'une série de n termes (la somme des valeurs absolues de ces écarts étant $= S_1$). Ici, encore, en vertu de la proposition 2, la somme de

---

(*) Nous nous sommes efforcés de déduire nos maxima et minima par des procédés élémentaires, car nous partageons pleinement l'opinion de M. LUCIEN MARCH qui attire l'attention sur le fait que « aujourd'hui la statistique pénètre dans la direction des affaires aussi bien que dans les sciences de la vie » et qu'il est donc intéressant que l'esprit de sa méthode soit aisément accessible « à ceux dont les connaissances en mathématiques sont rudimentaires, à la condition toutefois que les notations symboliques ou figurées, indispensables pour la clarté et la précision des raisonnements abstraits, ne les rebutent point ». (*Différences et corrélations en statistique* dans le *Journal de la Société de Statistique de Paris*, 1928). — M. CORRADO GINI s'exprime encore plus catégoriquement dans les termes lapidaires qui suivent; d'après lui, le mot d'ordre des statisticiens « instead of being represented by the words « Statistics with Mathematics », should be represented by « Statistics with the least mathematical means possible » (*Contribution of Italy to Modern Statistical Methods* dans *The Journal of the R. Statistical Society*, London 1926).

leurs carrés ($\Sigma\, d^2$) atteindra son *minimum* lorsqu'elles seront toutes égales entre elles ; elle sera alors égale à :

$$\sum_{\min} d^2 = \frac{(\Sigma\, d)^2}{n} \left(= \frac{S_1^2}{n} \right) \qquad [45]$$

Si nous désignons les valeurs absolues des écarts accusés par chacun d'eux par rapport à leur propre moyenne arithmétique (d_m) par δ (écarts moyennaux du second ordre), nous pouvons, d'après la formule [42], écrire que la somme des carrés de nos écarts du premier ordre est égale à :

$$\Sigma\, d^2 = \frac{S_1^2}{n} + \Sigma\, \delta^2 \qquad [46]$$

Elle augmentera avec l'augmentation des δ, c'est-à-dire avec les écarts du second ordre ou, en d'autres termes, avec l'accroissement de l'inégalité entre nos écarts du premier ordre.

Jusqu'ici, nos déductions générales s'appliquent entièrement à ce cas particulier. Mais la somme des carrés des écarts du premier ordre ($\Sigma\, d^2$) peut-elle atteindre le *maximum* de S_1^2 ? — Non. Car ce maximum n'est atteint que lorsque tous les variables sauf un deviennent $= 0$, le seul variable resté valable devenant à lui seul $= S_1$ (proposition 1). Or, il n'est pas possible qu'il n'y ait qu'un seul écart à la moyenne ; car dans toute série un écart positif ou négatif de la moyenne suppose forcément au moins *un* écart en sens opposé (autrement, la somme algébrique des écarts ne serait pas $= 0$ et la « moyenne » ne serait pas une moyenne). Quand les variables envisagées sont des écarts moyennaux, il faut donc qu'au moins *une* de ces variables soit positive et *une* négative, c'est-à-dire qu'au moins *deux* des variables ne soient pas égales à 0.

Pour établir quel est le maximum possible de $\Sigma\, d^2$, envisageons séparément les valeurs absolues des écarts positifs (que nous désignerons par x) et celles des écarts négatifs (que nous désignerons par z).

Remarquons d'abord que la somme des écarts moyennaux positifs étant dans toute série égale à la somme des valeurs absolues des écarts négatifs, nous pouvons écrire :

$$\Sigma\, x = \Sigma\, z = \frac{\Sigma\, d}{2} = \frac{S_1}{2}; \qquad (1)$$

nous avons en outre :

$$\Sigma\, d^2 = \Sigma\, x^2 + \Sigma\, z^2 \qquad (2)$$

D'un autre côté, les variations des x (à l'intérieur de leur somme $\frac{S_1}{2}$ qui demeure constante) et celles de z sont indépendantes les unes des autres (les écarts positifs peuvent être, par exemple, relativement nombreux et petits tandis que les écarts négatifs seraient peu nombreux et grands, et ainsi de suite). Dans ces conditions, pour que $\Sigma\, d^2$ devienne maxima, il faut que chacun des deux termes indépendants de la somme $(\Sigma\, x^2 + \Sigma\, z^2)$ devienne maximum, c'est-à-dire :

$$\underset{\max}{\Sigma\, d^2} = \underset{\max}{\Sigma\, x^2} + \underset{\max}{\Sigma\, z^2} \tag{3}$$

Mais d'après la proposition 1 du § précédent, $\Sigma\, x^2$ deviendra maximum quand tous les x sauf un deviennent $= 0$, le seul x resté valable étant égal à la somme de tous les x, soit $= \frac{S_1}{2}$; la somme des carrés des écarts positifs $(\Sigma\, x^2)$ est alors égale à $\left(\frac{S_1}{2}\right)^2$. Il en est exactement de même de $\Sigma\, z^2$ qui devient aussi $= \left(\frac{S_1}{2}\right)^2$. Notre dernière équation (3) peut donc être écrite comme suit :

$$\underset{\max}{\Sigma\, d^2} = \left(\frac{S_1}{2}\right)^2 + \left(\frac{S_1}{2}\right)^2 = \frac{S_1^2}{2}. \tag{47}$$

Nous avons ainsi les deux limites des variations de $\Sigma\, d^2$, et nous pouvons écrire, en réunissant les formules [45] et [47] :

$$\frac{S_1^2}{2} \geqq \Sigma\, d^2 \geqq \frac{S_1^2}{n} \tag{48}$$

ou, ce qui revient au même :

$$\frac{(\Sigma\, d)^2}{2} \geqq \Sigma\, d^2 \geqq \frac{(\Sigma\, d)^2}{n} \tag{48bis}$$

En d'autres termes : la somme des carrés des écarts moyennaux $(\Sigma\, d^2)$ d'une série (dont le nombre et la somme des termes ainsi que la somme des écarts sont constants) devient minima quand tous les écarts sont égaux entre eux, c'est-à-dire [46] quand les écarts moyennaux du second ordre (δ) sont tous $= 0$; elle est alors égale à $\frac{(\Sigma\, d)^2}{n}$;

la somme des carrés des écarts moyennaux (du premier ordre) augmente avec l'augmentation des écarts du second ordre [46], c'est-à-dire avec l'inégalité entre les écarts du premier ordre, et atteint son *maximum quand il n'y a qu'un seul écart positif et un seul écart négatif*, tous les autres écarts devenus $= 0$; elle est alors $= \frac{(\Sigma d)^2}{2}$. Si nous nous rappelons les définitions données au § 15, nous pouvons donc dire que lorsque la somme et le nombre des termes ainsi que la somme des écarts sont constants *la somme des carrés des écarts moyennaux* (Σd^2) *est minima quand la concentration des variations est nulle, que cette somme augmente avec l'augmentation de la concentration et qu'elle atteint son maximum quand la concentration est maxima, c'est-à-dire quand il n'y a qu'un seul écart moyennal positif et un seul négatif.* — Le maximum de Σd^2 correspond à la courbe A de notre graphique III (§ 15) ; le minimum de Σd^2 correspond à la courbe B du même graphique. — Le nombre (n) et la somme des valeurs absolues des écarts moyennaux ($\Sigma d = S_1$) étant constants, la somme des carrés de ces écarts peut être considérée comme l'expression propre (et dont les limites nous sont maintenant connues) de la concentration de la série des variations.

§ 46. — Extrayons maintenant la racine carrée de tous les termes de notre expression [48] ou [48^bis]. Nous aurons alors :

$$\frac{\Sigma d}{\sqrt{2}} \geqq \sqrt{\Sigma d^2} \geqq \frac{\Sigma d}{\sqrt{n}} .$$

Divisons encore tous les termes de cette expression par la somme des valeurs absolues des écarts moyennaux (Σd) afin de rendre les limites des variations indépendantes de celle-ci. Nous trouvons alors :

$$\frac{1}{\sqrt{2}} \geqq \frac{\sqrt{\Sigma d^2}}{\Sigma d} \geqq \frac{1}{\sqrt{n}} ,$$

soit, en multipliant tous les termes de cette expression par $\sqrt{n}$:

$$\sqrt{\frac{n}{2}} \geqq \sqrt{\frac{n \, \Sigma d^2}{\Sigma^2 d}} \geqq 1 . \qquad [49]$$

Mais le terme du milieu $\sqrt{\frac{n \, \Sigma d^2}{\Sigma^2 d}} = \frac{\sigma}{d_m}$ [19] ; nous pouvons donc

écrire :

$$\sqrt{\frac{n}{2}} \overline{>} \frac{\sigma}{d_m} \overline{>} 1 . \qquad [50]$$

Ainsi donc, *dans toute série de variations, le rapport de l'écart quadratique moyen* (σ) *à l'écart arithmétique moyen* (d_m) *varie avec la concentration des écarts de la série entre* 1 *au minimum et* $\sqrt{\frac{n}{2}}$ *au maximum*, n étant le nombre des termes de la série. — Nous voyons maintenant plus nettement la nature et les limites de notre indice de concentration σ_d (§ 33) (*).

Si nous prenons les carrés des termes de l'expression [49], nous avons

$$\frac{n}{2} \overline{>} \frac{n \sum d^2}{\sum^2 d} \overline{>} 1 . \qquad [52]$$

Le terme du milieu $\frac{n \sum d^2}{\sum^2 d}$ est notre indice de concentration des écarts C_n (§ 34). Nous pouvons donc écrire :

$$\frac{n}{2} \overline{>} C_n \overline{>} 1 . \qquad [53]$$

Nous sommes ainsi fixés sur les variations de C_n : il est à son minimum 1 quand la concentration des écarts est nulle (tous les écarts moyennaux étant égaux), il monte avec leur concentration (avec les écarts moyennaux du second ordre, soit avec l'inégalité des écarts du premier ordre) et atteint le maximum de $\frac{n}{2}$ quand la concentration est maxima, c'est-à-dire quand la série n'accuse qu'un seul écart positif et un seul négatif.

§ 47. — Reprenons maintenant notre formule [50] et enlevons à chacun de ses termes le nombre 1. Nous avons alors :

$$\sqrt{\frac{n}{2}} - 1 > \left(\frac{\sigma}{d_m} - 1\right) \overline{>} 0 . \qquad [54]$$

(*) Il suit en outre de la formule [50] que : $\sqrt{\frac{n}{2}}\, d_m \overline{>} \sigma \overline{>} d_m$. [51]

On voit ainsi que ce qui varie avec la concentration plus ou moins grande des écarts de la série, ce n'est pas à proprement parler la quantité $\frac{\sigma}{d_m}$, mais bien plutôt $\left(\frac{\sigma}{d_m} - 1\right)$: c'est cette dernière expression qui varie et disparaît avec les variations et la disparition de la concentration ; l'indice $\frac{\sigma}{d_m}$ est au fond la somme de deux éléments : 1 et $\left(\frac{\sigma}{d_m} - 1\right)$, dont l'un (1) est constant, indépendant de la mesure de la concentration des écarts et subsiste même quand la concentration est nulle, et dont l'autre $\left(\frac{\sigma}{d_m} - 1\right)$ varie seul avec la concentration et disparaît avec elle. C'est donc à proprement parler $\left(\frac{\sigma}{d_m} - 1\right)$ qui marque la vraie mesure de la concentration des écarts ; elle varie entre 0, quand la concentration des écarts est nulle et $\left(\sqrt{\frac{n}{2}} - 1\right)$ quand la concentration des écarts est à son maximum.

Divisant maintenant tous les membres de l'expression [54] par son membre gauche, nous trouvons :

$$1 \geqq \frac{\frac{\sigma}{d_m} - 1}{\sqrt{\frac{n}{2}} - 1} \geqq 0 \qquad [55]$$

Le terme du milieu de cette expression qui montre l'importance d'une concentration donnée $\left(\frac{\sigma}{d_m} - 1\right)$ *par rapport au maximum de concentration possible* $\left(\sqrt{\frac{n}{2}} - 1\right)$ présente ainsi une fraction positive, variant entre 0 et 1 et marquant de la sorte très nettement (par des fractions pouvant être 0,10, 0,40, 0,70, etc.) tous les *degrés* de concentration possibles : c'est un réel *coefficient de la concentration des*

écarts. Nous le désignerons désormais par c_n (*). Nous pouvons donc écrire :

$$c_n = \frac{\dfrac{\sigma}{d_m} - 1}{\sqrt{\dfrac{n}{2} - 1}} \,. \qquad [56]$$

Si l'on substitue, dans cette formule, à $\dfrac{\sigma}{d_m}$ sa valeur $\sqrt{\dfrac{n \,\Sigma\, d^2}{\Sigma^2\, d}}$, notre coefficient de concentration c_n prend la forme développée (qui montre toutes les opérations nécessaires à son calcul) que voici :

$$c_n = \frac{\sqrt{\dfrac{n \,\Sigma\, d^2}{\Sigma^2\, d}} - 1}{\sqrt{\dfrac{n}{2} - 1}} \; (**) \qquad [57]$$

(*) Dans notre étude *Les fluctuations saisonnières du chômage dans l'industrie du bâtiment* (Genève, 1929, « Bureau International du Travail ») où nous avons pour la première fois fait un usage systématique de ce coefficient, nous l'avions désigné par la lettre β. Nous préférons l'appeler c.

(**) En raisonnant de façon analogue, on pourrait déduire un coefficient de concentration, variant entre 0 et 1, également sur la base de l'indice C_n qui est égal à $\left(\dfrac{\sigma}{d_m}\right)^2$; ce coefficient que nous dirons c' serait égal à:

$$c'_n = \frac{C_n - 1}{\dfrac{n}{2} - 1} \qquad [58]$$

Son calcul serait même plus simple que celui de c_n, mais son expression serait moins nette. Car [d'après 50] $\dfrac{\sigma}{d_m} < \sqrt{\dfrac{n}{2}}$; en prenant le carré de ces deux termes (dont le premier, le plus petit, entre dans le numérateur du coefficient et le second, le plus grand, dans son dénominateur, et qui sont > 1), nous augmentons le dénominateur plus que le numérateur, nous réduisons ainsi le résultat final et le rendons moins expressif. C'est ici juste le contraire de ce que nous avons vu plus haut pour les indices inachevés C et $\dfrac{\sigma}{d_m}$ Dans la suite nous nous servirons exclusivement du coefficient c_n.

§ 48. — Le coefficient de la concentration des écarts, déduit sans aucune condition particulière qui soit propre aux seules séries périodiques, est valable pour toutes sortes de séries, périodiques ou non. Pour les séries périodiques les plus fréquentes, comprenant de beaucoup le plus souvent 12, 24 ou 7 termes, la formule du coefficient c_n peut être sensiblement simplifiée. En effet, substituant à n successivement les nombres 12, 24 et 7 dans les formules [56] et [57], on trouve :

Séries à 12 termes : à la base de $\frac{\sigma}{d_m}$ [56] — à la base de $\frac{\Sigma d^2}{\Sigma^2 d}$ [57]

$$c_{12} = \frac{\frac{\sigma}{d_m} - 1}{\sqrt{6} - 1} = \frac{\frac{\sigma}{d_m} - 1}{2,45 - 1} = 0,69 \left(\frac{\sigma}{d_m} - 1 \right) = 0,69 \left(\sqrt{\frac{12 \, \Sigma d^2}{\Sigma^2 d}} - 1 \right) \qquad [59]$$

Séries à 24 termes :

$$c_{24} = \frac{\frac{\sigma}{d_m} - 1}{\sqrt{12} - 1} = \frac{\frac{\sigma}{d_m} - 1}{2,464} = 0,406 \left(\frac{\sigma}{d_m} - 1 \right) = 0,406 \left(\sqrt{\frac{24 \, \Sigma d^2}{\Sigma^2 d}} - 1 \right) \qquad [60]$$

Séries à 7 termes :

$$c_7 = \frac{\frac{\sigma}{d_m} - 1}{\sqrt{3,5} - 1} = \frac{\frac{\sigma}{d_m} - 1}{0,871} = 1,148 \left(\frac{\sigma}{d_m} - 1 \right) = 1,148 \left(\sqrt{\frac{7 \, \Sigma d^2}{\Sigma^2 d}} - 1 \right) \qquad [61]$$

Selon que l'on aura déjà calculé σ et d_m pour d'autres besoins ou non, on se servira de préférence des formules à base de $\frac{\sigma}{d_m}$ ou de celles à base de $\frac{\Sigma d^2}{\Sigma^2 d}$. Enfin, pour faciliter davantage l'application de notre coefficient de concentration c_n, nous avons calculé la valeur de c_n pour toutes les valeurs possibles de $\frac{\sigma}{d_m}$ et de $\frac{\Sigma d^2}{\Sigma^2 d}$ à 0,01 près, et cela pour les séries à 12 termes (c_{12}), à 24 termes (c_{24}) et à 7 termes (c_7). Celui qui voudrait en faire usage trouvera les tables des valeurs de c_n comme annexe à la présente étude (Annexe I, tables I, *A* et I, *B*).

Nos dernières formules [59 à 61] montrent en outre très clairement que pour des séries homonomes les *augmentations* (et les *diminutions*) du coefficient c_n sont *proportionnelles* aux augmentations (et diminutions) de l'indice de concentration $\frac{\sigma}{d_m}$. Ainsi, pour c_{12}, à toute augmentation (ou diminution) de 0,1, 0,2, 1,0, etc., de $\frac{\sigma}{d_m}$ correspond une augmentation (ou diminution) de $0{,}69 \times 0{,}1$, $0{,}69 \times 0{,}2$, $0{,}69 \times 1{,}0$, etc., de c_{12}. De même, dans les séries à 24 termes, à toute augmentation de $\frac{\sigma}{d_m}$ correspond pour c_{24}, la même augmentation multipliée par 0,406, et ainsi de suite. — D'une façon générale, d'après [56], *à toute augmentation (ou diminution) de* $\frac{\sigma}{d_m}$ *correspond la même augmentation (ou diminution) multipliée par* $\frac{1}{\sqrt{\frac{n}{2}-1}}$, *du coefficient de la concentration des écarts* c_n.

§ 49. — Mais puisque $\frac{\sigma}{d_m}$ augmente avec la concentration de la série des écarts (n étant constant), son inverse $\frac{d_m}{\sigma}$ diminue avec l'augmentation de la concentration, atteint son minimum quand $\frac{\sigma}{d_m}$ (c'est-à-dire la concentration) est à son maximum et devient maximum quand la concentration est minima; $\frac{d_m}{\sigma}$ devient ainsi l'indice (inachevé) du caractère opposé à la concentration que nous avons appelé (§ 15) le nivellement (ou l'extension) des variations. Il en est exactement de même de l'inverse de C_n. Au lieu d'envisager les indices de la concentration $\frac{\sigma}{d_m}$ ou C_n, nous pouvons donc prendre aussi leurs inverses $\frac{d_m}{\sigma}$ ou $\frac{1}{C_n} = \frac{\Sigma^2 d}{n \, \Sigma d^2}$ comme *indices du nivellement* des variations.

Raisonnant ensuite d'une façon parfaitement analogue à celle exposée aux §§ 45 et 46, nous pouvons dégager également le *coefficient du nivellement* des variations (périodiques ou non).

En effet, d'après [50], nous pouvons écrire :

$$1 \geqq \frac{d_m}{\sigma} \geqq \sqrt{\frac{2}{n}}\,; \tag{62}$$

d'où :

$$1 - \sqrt{\frac{2}{n}} \geqq \left(\frac{d_m}{\sigma} - \sqrt{\frac{2}{n}}\right) \geqq 0$$

soit :

$$1 \geqq \frac{\dfrac{d_m}{\sigma} - \sqrt{\dfrac{2}{n}}}{1 - \sqrt{\dfrac{2}{n}}} \geqq 0\,. \tag{63}$$

Le terme du milieu marque le degré du nivellement des variations d'une série donnée par rapport au nivellement maximum possible et varie entre 0 (nivellement minimum) et 1 (nivellement parfait) ; c'est le *coefficient du nivellement des variations* que nous désignerons par un ν grec. Nous pouvons ainsi écrire :

$$\nu_n = \frac{\dfrac{d_m}{\sigma} - \sqrt{\dfrac{2}{n}}}{1 - \sqrt{\dfrac{2}{n}}} = \frac{\sqrt{\dfrac{\Sigma^2 d}{n \Sigma d^2}} - \sqrt{\dfrac{2}{n}}}{1 - \sqrt{\dfrac{2}{n}}}\,. \tag{64}$$

Appliqué aux séries de 12, de 24 et de 7 termes, ce coefficient de nivellement devient respectivement :

$$\nu_{12} = \frac{\dfrac{d_m}{\sigma} - \sqrt{\dfrac{1}{6}}}{1 - \sqrt{\dfrac{1}{6}}} = \frac{\dfrac{d_m}{\sigma} - 0{,}408}{1 - 0{,}408} = \frac{\dfrac{d_m}{\sigma} - 0{,}408}{0{,}592} =$$

$$= 1{,}69 \left(\frac{d_m}{\sigma} - 0{,}408\right) (*) = 1{,}69 \left(\sqrt{\frac{\Sigma^2 d}{12 \Sigma d^2}} - 0{,}408\right) \tag{65}$$

(*) Dans notre étude citée (*Fluctuations saisonnières du chômage*, etc., § 16) une faute typographique a mis dans cette expression le facteur 0,69 (au lieu de 1,69) et dans la formule du coefficient de la concentration au

$$\nu_{24} = \frac{\frac{d_m}{\sigma} - \sqrt{\frac{1}{12}}}{1 - \sqrt{\frac{1}{12}}} = \frac{\frac{d_m}{\sigma} - 0{,}289}{0{,}711} = 1{,}406\left(\frac{d_m}{\sigma} - 0{,}289\right) =$$

$$= 1{,}406\left(\sqrt{\frac{\Sigma^2 d}{24\,\Sigma\, d^2}} - 0{,}289\right) \qquad [66]$$

$$\nu_7 = \frac{\frac{d_m}{\sigma} - \sqrt{\frac{2}{7}}}{1 - \sqrt{\frac{2}{7}}} = \frac{\frac{d_m}{\sigma} - 0{,}535}{1 - 0{,}535} = 2{,}15\left(\frac{d_m}{\sigma} - 0{,}535\right) =$$

$$= 2{,}15\left(\sqrt{\frac{\Sigma^2 d}{7\,\Sigma\, d^2}} - 0{,}535\right). \qquad [67]$$

Les tables II, *A* et II, *B* (Annexe I) donnent directement les coefficients de nivellement (ν) en fonction de $\frac{d_m}{\sigma}$ et de $\frac{\Sigma^2 d}{\Sigma\, d^2}$ pour les séries à 12, à 24 et à 7 termes.

Comme on le voit aisément d'après nos dernières formules, les augmentations (ou diminutions) du coefficient de nivellement (ν) sont pour des séries homonomes *proportionnelles* aux augmentations (ou diminutions) de $\frac{d_m}{\sigma}$; à chaque augmentation (ou diminution) de $\frac{d_m}{\sigma}$ correspond la même augmentation (ou diminution) multipliée par $\frac{1}{1 - \sqrt{\frac{2}{n}}}$, du coefficient de nivellement ν_n.

Notons encore que les augmentations (ou diminutions) de $\frac{d_m}{\sigma}$ n'étant pas proportionnelles à celles de son inverse $\frac{\sigma}{d_m}$, les augmen-

---

contraire 1,69 (à la place de 0,69). Dans la même étude, nous avons employé le terme de *prolongation* des variations pour ce que nous appelons ici *nivellement* (ou extension) et le coefficient de ce caractère avait été désigné par γ (au lieu de ν).

tations (ou diminutions) du coefficient de nivellement (ν_n) ne sont généralement ni égales ni proportionnelles aux diminutions (ou augmentations) du coefficient de concentration (c_n), et vice versa.

§ 50. — Afin d'éviter toute confusion, montrons explicitement que, pour une série donnée, *nos coefficients de concentration* (c_n) *et de nivellement* (ν_n) *ne sont complémentaires*, c'est-à-dire que leur somme n'est égale à 1, *que dans les cas limites où l'un des deux est égal à* 1 *et l'autre à* 0.

En effet, pour simplifier les opérations, désignons $\frac{\sigma}{d_m}$ par x et $\sqrt{\frac{n}{2}}$ par y; $\frac{d_m}{\sigma}$ devient alors $\frac{1}{x}$ et $\sqrt{\frac{2}{n}}$ devient $= \frac{1}{y}$; la somme des deux coefficients c et ν devient alors :

$$c_n + \nu_n = \frac{x-1}{y-1} + \frac{\frac{1}{x} - \frac{1}{y}}{1 - \frac{1}{y}} = \frac{x-1}{y-1} + \frac{(y-x)\,y}{x\,y\,(y-1)} =$$

$$= \frac{x\,(x-1) + y - x}{x\,(y-1)} = \frac{x^2 - 2x + y}{x\,y - x} \quad \ldots\ldots (1)$$

Pour que cette expression soit $= 1$, il faut que le numérateur soit égal au dénominateur, soit :

$$\begin{aligned} & x^2 - 2x + y = x\,y - x\,; \\ & x^2 - x + y - x\,y = 0\,; \\ & x\ (x-1) - y\ (x-1) = 0\,; \\ & (x-1)\ (x-y) = 0\,. \qquad\qquad (2) \end{aligned}$$

Pour que cette dernière expression soit $= 0$, il faut ou bien que $x = 1$ ou que $x = y$. Mais x, c'est-à-dire $\frac{\sigma}{d_m}$, est $= 1$ dans le cas de concentration nulle [50], soit de nivellement maximum ; x est $= y$, c'est-à-dire $\frac{\sigma}{d_m} = \sqrt{\frac{n}{2}}$, dans le cas de concentration maxima [50] soit de nivellement nul. — C. Q. F. D.

D'un autre côté, toujours d'après [50], $y \gtreqless x \gtreqless 1$; à l'exception des deux cas extrêmes dont il vient d'être question, on a donc : $y > x > 1$; le produit $(x - 1)\ (x - y)$, dont l'un des facteurs $(x - 1)$ est positif (puisque $x > 1$) et l'autre $(x - y)$ est négatif (puisque $y > x$), est donc négatif, soit :

$$(x - 1)\ (x - y) < 0 ;$$

$$x^2 - x - xy + y < 0 ;$$

$$x^2 - 2x + y < xy - x ;$$

il s'ensuit donc que : $\dfrac{x^2 - 2x + y}{xy - x} < 1,$

soit, d'après la formule (1) du présent § :

$$(c_n + \nu_n) < 1 \qquad [68]$$

Donc, à l'exception des deux cas extrêmes de concentration maxima et de concentration nulle (ou nivellement nul et nivellement parfait), la somme des deux coefficients est inférieure à 1.

§ 51. — Nous avons vu plus haut que les indices inachevés (σ_d , C_n , etc.) ont le grand désavantage de ne jamais marquer le degré du caractère considéré, mais de montrer seulement dans quelle série ce caractère est plus accentué et dans laquelle il l'est moins. Précisons maintenant l'autre restriction à l'emploi légitime des indices inachevés de concentration $\frac{\sigma}{d_m}$ et C_n, la limitation de leur application aux seules séries homonomes. En effet, d'après [53], nous voyons que le maximum de C_n est égal à $\frac{n}{2}$; de même [50] le maximum de $\frac{\sigma}{d_m}$ est $= \sqrt{\frac{n}{2}}$; le maximum des indices inachevés de la concentration varie donc avec n, avec le nombre des termes des séries envisagées (le maximum de C_n varie proportionnellement à ce nombre, le maximum de $\frac{\sigma}{d_m}$ varie proportionnellement aux racines carrées de ce nombre). Par conséquent, la même valeur numérique de C_n ou de $\frac{\sigma}{d_m}$ qui signifie pour une série une très forte concentration signifiera pour une autre série ayant

plus de termes une concentration bien inférieure, et vice versa. Ainsi, par exemple, $C_{12} = 6 \left(\text{ou } \frac{\sigma}{d_m} = 2{,}45\right)$ marquerait le maximum de concentration pour une série à 12 termes $\left(\text{maximum de } C_n = \frac{n}{2} = \frac{12}{2} = 6 \text{ ; maximum de } \frac{\sigma}{d_m} = \sqrt{\frac{n}{2}} = \sqrt{6} = 2{,}45\right)$ tandis que C_{24} égal même à 8 ou 9 ou 10 $\left(\text{soit } \frac{\sigma}{d_m} = 2{,}8 \text{ ou } 3 \text{ ou } 3{,}2\right)$ marquerait une concentration sensiblement au-dessous du maximum pour une série à 24 termes $\left(\text{pour une telle série, maximum de } C_{24} = \frac{24}{2} = 12 \text{ et maximum de } \frac{\sigma}{d_m} = \sqrt{12} = 3{,}46\right)$. Il est donc clair que les indices inachevés ne peuvent nous montrer laquelle des séries est plus concentrée et laquelle l'est moins que lorsqu'il s'agit de séries homonomes (n demeurant constant). Appliqués à des séries hétéronomes, ces indices inachevés peuvent nous induire en erreur (*).

Il en est autrement du *coefficient* de la concentration (c_n) qui, lui, nous montre toujours le degré de concentration d'une série par rapport au maximum de concentration possible pour une série ayant son nombre de termes : la différence des maxima possibles provenant de la différence du nombre des termes est déjà prise en considération par les coefficients : la comparaison devient par suite légitime même entre séries hétéronomes (**).

(*) Voici encore un exemple schématique illustrant cette propriété: Soient A' et A'' deux séries hétéronomes accusant les écarts moyennaux suivants :

A' . . . . . $+6, 0, -6, 0$

A'' . . . . . $+6, +6, +6, 0, 0, 0, -6, -6, -6, 0, 0, 0$.

Dans les deux cas on aurait le même $\sigma = \sqrt{18}$ et le même $d_m = 3$, et pourtant l'une des séries (A') aurait une concentration maxima des écarts moyennaux et l'autre (A'') une concentration bien médiocre.

(**) Pour les deux séries hétéronomes A' et A'' citées à titre d'illustration dans la note précédente et ayant, toutes les deux, l'indice de concentration $\frac{\sigma}{d_m} = \frac{\sqrt{18}}{3} = \sqrt{2} = 1{,}41$, nous trouvons les coefficients de concentration que

§ 52. — Il faut cependant noter que la comparaison de la concentration de séries hétéronomes n'est pas toujours légitime, indépendamment même du procédé de cette comparaison.

Lorsqu'il s'agit de séries périodiques, il y a lieu de distinguer sous ce rapport deux modalités : 1° la différence du nombre des termes vient de la nature même des périodicités comparées, et 2° cette différence est arbitraire, résultant simplement du fait que le nombre des observations par période opérées par nous a été dans un cas plus petit ou plus grand que dans d'autres. Exemples de comparaisons de la première espèce : la concentration des dépôts effectués dans les caisses d'épargne selon les jours (ouvrables) de la semaine (série à 6 termes). selon les jours du mois (série à 26 termes) et selon les mois de l'année (série à 12 termes) ; ou bien la concentration des suicides selon les jours de la semaine (7 termes) et selon les mois de l'année (12 termes). Exemples de la seconde espèce : concentration des variations de la température du corps humain au cours de la journée selon qu'on l'observe toutes les heures (24 termes) ou une fois par deux heures (12 termes) ; ou bien la concentration de la périodicité annuelle des variations de la température de l'air atmosphérique observée une fois par mois et celle de la périodicité journalière de la température de l'air observée toutes les heures (24 termes), et ainsi de suite.

Le nombre des termes dans la périodicité hebdomadaire des dépôts (6) ou des suicides (7) n'est pas chose arbitraire, la journée constituant une unité naturelle (plus exactement, une période fermée d'un ordre plus restreint) ; il en est de même du nombre des jours dans le mois ou des mois dans l'année quand il s'agit de phénomènes économiques et sociaux, le mois constituant pour bien des phénomènes, de cette nature (traitements, salaires, pensions, règlements de comptes,

---

voici :

$$A' \ldots c_4 = \frac{\sqrt{2} - 1}{\sqrt{2} - 1} = 1 \;;\; A'' \ldots c_{12} = \frac{\sqrt{2} - 1}{\sqrt{6} - 1} = 0{,}69 \,.\, 0{,}41 = 0{,}28\,.$$

Quelle différence !

Pour remédier au défaut des indices inachevés que nous venons de signaler, on pourrait les diviser par leur maximum, c'est-à-dire qu'on pourrait prendre $\frac{C_n}{n:2}$ ou $\frac{\sigma}{d_m} : \sqrt{\frac{n}{2}}$; mais ces expressions n'auraient ni la simplicité des indices inachevés C_n et $\frac{\sigma}{d_m}$ ni la netteté du coefficient c_n.

loyers, etc.) une véritable unité. La différence du nombre des termes est ici une différence *objective* tenant à la nature des périodicités examinées. Il en est autrement du nombre des observations de la température au cours de l'année (ou au cours de la journée) : on peut en faire par exemple 4 (une fois par trimestre), 12 (par mois), 24 (par quinzaine), etc. ; tout ce qu'on peut dire ici c'est qu'en prenant 12 mensurations (chaque mois, respectivement toutes les deux heures), on serre la périodicité réelle de plus près qu'en en prenant seulement 4 (une fois par trimestre, resp. par 6 heures) et qu'en prenant 24 mensurations (chaque quinzaine, resp. toutes les heures) nous la serrons de plus près encore, nous la précisons davantage. Mais dans ce cas, comment comparer la concentration de deux périodicités observées avec des précisions différentes ? Supposons, en effet, qu'on a fait 4 observations (chaque trimestre par exemple) de la périodicité A_1 et 12 de la périodicité A_2 et qu'on a trouvé les écarts moyennaux suivants:

A_1 .... 0, + 6 0 — 6

A_2 .... 0, 0, 0, + 5, + 6, + 4, 0, 0, 0, — 6, — 6, — 3

Qu'est-ce que cela signifie ? Dans la série à 4 termes (A_1) nous trouvons une concentration maxima ; elle est loin du maximum dans la série (A_2). Si A_1 et A_2 se rapportent à la même périodicité (par exemple, à la périodicité annuelle de la température atmosphérique d'un endroit donné), la série A_2 comporte simplement la *rectification* de la série A_1, montrant que la concentration des variations, qui a pu *paraître* maxima d'après A_1, est au contraire dans la réalité assez faible. Dans des cas pareils, c'est la série qui a plus de termes qui serre la réalité de plus près et l'image donnée par la série ayant moins de termes doit être tout simplement abandonnée. — Si A_1 et A_2 se rapportent à des périodicités différentes (par exemple, périodicité journalière de la température observée toutes les six heures et périodicité annuelle de la température observée tous les mois) ou à la même périodicité observée pour des phénomènes différents (par exemple, périodicité annuelle de la température observée par mois et celle du chômage observé par trimestre, ou encore périodicité journalière de la température du corps humains observée à un endroit, chez un sexe, groupe sanguin, etc., 4 fois par jour et dans un autre endroit, sexe, groupe sanguin, etc., 12 fois par jour), alors aucune comparaison des deux séries n'est possible ; car nous ne savons rien sur la nature de la série que l'on aurait pour le phénomène A_1 si on l'avait soumis également à 12 observations par période.

Quand le nombre des termes des séries périodiques est arbitraire, la comparaison de ces séries n'a donc de sens et n'est relativement légitime (dans les limites des observations faites) que s'il s'agit de séries homonomes. La comparaison des séries hétéronomes demeure pourtant légitime et instructive quand la différence du nombre des termes découle de la nature même des périodicités envisagées.

Nous avons examiné ici la comparabilité de séries périodiques hétéronomes au point de vue de leur concentration (ou nivellement). Mais les mêmes observations et les mêmes distinctions doivent être faites aussi au sujet de la plupart des autres caractères des séries périodiques, notamment concernant le nombre de leurs sommets, la situation de ceux-ci dans la série, la précipitation ou gradation des séries, leur symétrie et dissymétrie, etc.

§ 53. — Nous avons vu (§ 51) que pour des séries hétéronomes les mêmes valeurs de $\frac{\sigma}{d_m}$ correspondent à des concentrations différentes. Il en est naturellement de même de l'inverse de $\frac{\sigma}{d_m}$, c'est-à-dire de l'indice du nivellement $\frac{d_m}{\sigma}$. D'un autre côté, nous avons vu que sauf dans les cas limites, nos coefficients de concentration (c_n) et de nivellement (ν_n) ne sont pas complémentaires ; à une augmentation donnée de c_n ne correspond donc pas forcément une diminution identique de ν_n, et vice versa. Dans ces conditions, il ne sera peut-être pas sans intérêt d'examiner encore quelles sont, pour des séries ayant les mêmes $\frac{\sigma}{d_m}$, mais différant par le nombre de leurs termes, les différences de la concentration et du nivellement mesurées par leurs coefficients respectifs (c_n et ν_n).

Prenons donc deux séries hétéronomes dont l'une a n' termes et l'autre n'' termes, mais qui ont le même écart moyennal quadratique moyen (σ) et le même écart moyennal arithmétique moyen (d_m), qui ont ainsi les mêmes indices inachevés de la concentration $\left(\frac{\sigma}{d_m}\right)$ et du nivellement $\left(\frac{d_m}{\sigma}\right)$. Quelle sera la différence des coefficients de concentration des deux séries ($c' - c''$) et quelle sera la différence de leurs coefficients de nivellement ($\nu' - \nu''$) ?

D'après les formules [56] et [64], nous avons :

$$c' - c'' = \frac{\frac{\sigma}{d_m} - 1}{\sqrt{\frac{n'}{2}} - 1} - \frac{\frac{\sigma}{d_m} - 1}{\sqrt{\frac{n''}{2}} - 1}\;;\quad \nu' - \nu'' = \frac{\frac{d_m}{\sigma} - \sqrt{\frac{2}{n'}}}{1 - \sqrt{\frac{2}{n'}}} - \frac{\frac{d_m}{\sigma} - \sqrt{\frac{2}{n''}}}{1 - \sqrt{\frac{2}{n''}}}\;;$$

Désignant $\frac{\sigma}{d_m}$ par x, $\sqrt{\frac{n'}{2}}$ par y et $\sqrt{\frac{n''}{2}}$ par z, nous pouvons ainsi écrire :

$$c' - c'' = \frac{x-1}{y-1} - \frac{x-1}{z-1} = \frac{(x-1)\,(z-1-y+1)}{(y-1)\,(z-1)} =$$

$$= \frac{(x-1)\,(z-y)}{(y-1)\,(z-1)}\;; \qquad (1)$$

$$\nu' - \nu'' = \frac{\frac{1}{x} - \frac{1}{y}}{1 - \frac{1}{y}} - \frac{\frac{1}{x} - \frac{1}{z}}{1 - \frac{1}{z}} = \frac{\frac{y-x}{xy}}{\frac{y-1}{y}} - \frac{\frac{z-x}{xz}}{\frac{z-1}{z}} =$$

$$= \frac{y-x}{x\,(y-1)} - \frac{z-x}{x\,(z-1)} = \frac{(y-x)\,(z-1) - (z-x)\,(y-1)}{x\,(y-1)\,(z-1)} =$$

$$= \frac{xy - xz - y + z}{x\,(y-1)\,(z-1)} = \frac{(x-1)\,(y-z)}{x\,(y-1\,(z-1)}\;; \qquad (2)$$

Le rapport de ces deux différences est donc :

$$\frac{c' - c''}{\nu' - \nu''} = \frac{(x-1)\,(z-y)\,.\,x\,(y-1)\,(z-1)}{(y-1)\,(z-1)\,(x-1)\,(y-z)} = \frac{x\,(z-y)}{y-z} =$$

$$= -x = -\frac{\sigma}{d_m} \qquad [69]$$

En d'autres termes, *pour des séries ayant le même indice inachevé de la concentration* $\frac{\sigma}{d_m}$ $\left(\text{ou du nivellement } \frac{d_m}{\sigma}\right)$, *mais différant par le nombre des termes, la différence des coefficients de concentration est à la*

différence des coefficients de nivellement comme l'écart moyennal quadratique moyen (σ) *est à l'écart moyennal arithmétique moyen* (d_m) ou, ce qui revient au même, *leur rapport est égal à l'indice de concentration* $\left(\frac{\sigma}{d_m}\right)$ *des séries considérées (les deux différences ayant, en outre, des signes opposés).*

De cette proposition, on peut tirer deux conséquences :

1° Puisque (sauf dans le cas limite de concentration nulle) $\sigma > d_m$, il s'ensuit de la formule [69] que, prises en valeur absolue,

$$[c' - c''] > [\nu' - \nu''] \qquad [70]$$

c'est-à-dire qu'avec la différence du nombre des termes des séries envisagées $\left(\frac{\sigma}{d_m}\text{ étant le même}\right)$, le coefficient de la concentration des écarts diffère plus d'une série à l'autre que le coefficient du nivellement.

2° Pour des séries hétéronomes ayant le même $\frac{\sigma}{d_m}$, la différence de concentration ($c' - c''$) est absolument parlant moindre quand σ est rapproché de d_m ; mais alors c_n est rapproché de 0, et une petite différence absolue des coefficients de concentration peut alors signifier une très grande différence relative $\left(\frac{c' - c''}{c'}\right)$; par contre, la différence des coefficients du nivellement ($\nu' - \nu''$), déjà un peu plus réduite que celle des coefficients de la concentration, est relativement $\left(\frac{\nu' - \nu''}{\nu'}\right)$ encore d'autant plus faible que les coefficients de nivellement sont alors particulièrement élevés (se rapprochant de 1).

Les séries hétéronomes ayant le même $\frac{\sigma}{d_m}$ diffèrent ainsi absolument et relativement davantage par leurs coefficients de concentration que par ceux de nivellement.

§ 54. — Pour déterminer les limites des variations de notre *indice de précipitation* des séries monocéphales périodiques (P_n) on peut procéder d'une façon parfaitement analogue à celle exposée aux §§ 45 et 46 pour la fixation des limites de l'indice de concentration (C_n) ; les résultats auxquels on aboutirait seraient également tout-à-fait

analogues à ceux obtenus pour la concentration, avec cette seule différence qu'à la place des écarts moyennaux (d), on aurait partout des écarts successifs (e). Pour varier un peu la démonstration, on nous permettra cependant de suivre une voie partiellement différente.

Soit une série périodique monocéphale ayant n termes et une amplitude α ; désignons les écarts successifs positifs (c'est-à-dire en allant de v_{min} à v_{max}) par t et les écarts successifs négatifs (allant de v_{max} à v_{min}) par u ; admettons encore que le nombre des écarts successifs positifs est $= m$, celui des négatifs est $= (n - m)$. La somme des écarts successifs positifs, comme celle des négatifs, étant égale à α, nous pouvons écrire d'après [44] :

$$\alpha^2 \overline{>} \Sigma t^2 \overline{>} \frac{\alpha^2}{m} \qquad (1)$$

$$\alpha^2 \overline{>} \Sigma u^2 \overline{>} \frac{\alpha^2}{n - m} \qquad (2)$$

Additionnant les termes correspondants de ces deux expressions, on trouve :

$$2\,\alpha^2 \overline{>} (\Sigma t^2 + \Sigma u^2) \overline{>} \alpha^2 \left(\frac{1}{m} + \frac{1}{n - m}\right),$$

soit :

$$2\,\alpha^2 \overline{>} \Sigma e^2 \overline{>} \frac{\alpha^2\,(n - m + m)}{m\,(n - m)},$$

c'est-à-dire :

$$2\,\alpha^2 \overline{>} \Sigma e^2 \overline{>} \frac{n\,\alpha^2}{m\,(n - m)}. \qquad (3)$$

Mais le produit des deux facteurs positifs : $m\,(n - m)$ dont la somme n demeure constante devient maximum quand chacun d'eux est $= \frac{n}{2}$ (d'après § 44, proposition 3) ; l'expression $\frac{n\,\alpha^2}{m\,(n - m)}$ devient alors minima, et sera $= \frac{n\,\alpha^2}{\left(\frac{n}{2}\right)^2} = \frac{4\,\alpha^2}{n}$. Substituant cette expression à la limite minima de Σe^2 (c'est-à-dire au membre droit de notre dernière formule (3)), nous avons :

$$2\,\alpha^2 \overline{>} \Sigma e^2 \overline{>} \frac{4\,\alpha^2}{n}. \qquad [71]$$

Il s'ensuit $\left(\text{en divisant tous les termes par } \frac{4\,\alpha^2}{n}\right)$ que :

$$\frac{n}{2} \overline{>} \frac{n \sum e^2}{4\,\alpha^2} \overline{>} 1\,. \qquad [72]$$

Or, le terme du milieu est, d'après [26^{bis}], précisément notre indice de précipitation P_n ; notre formule [72] peut donc s'écrire encore sous la forme suivante :

$$\frac{n}{2} \overline{>} P_n \overline{>} 1\,. \qquad [73]$$

Les formules [72] et [73] sont la transposition aux écarts successifs et à la précipitation des séries périodiques monocéphales des formules [51] et [52] dégagées plus haut pour les écarts moyennaux et la concentration des séries. P_n, comme C_n, a pour minimum le chiffre 1 qui (d'après § 44, proposition 2) correspond au cas où tous les écarts successifs sont égaux ; P_n, comme C_n, augmente avec l'augmentation de l'inégalité entre les écarts (d'après la formule [42]), qui naturellement sont ici les écarts successifs ; enfin, comme C_n, P_n devient maximum quand il n'y a qu'un seul écart (successif) positif et un seul négatif, égaux chacun en valeur absolue à la demi-somme des écarts, c'est-à-dire ici $= \alpha$; P_n est alors, comme C_n, égal à $\frac{n}{2}$.

Si nous extrayons les racines carrées des termes de notre formule [72], et si nous remplaçons $\sqrt{\frac{n \sum e^2}{4\,\alpha^2}}$ par son équivalent $\frac{\varsigma}{e_m}$ d'après [25], nous pouvons écrire $\left(\text{comme pour } \frac{\sigma}{d_m}\right)$:

$$\sqrt{\frac{n}{2}} \overline{>} \frac{\varsigma}{e_m} \overline{>} 1\,. \qquad [74]$$

§ 55. — Nous avons ainsi les deux limites de nos indices inachevés de la précipitation. Raisonnant comme au § 47, nous pouvons maintenant dégager le *coefficient de la précipitation* des séries périodiques monocéphales montrant le degré de précipitation d'une série donnée par rapport au maximum de précipitation possible pour une série

de la même espèce. Nous le désignerons par p_n. Il résulte en effet de la formule [74] que :

$$1 \geqq \frac{\frac{\varsigma}{e_m} - 1}{\sqrt{\frac{n}{2} - 1}} \geqq 0\,. \qquad [75]$$

Le coefficient cherché est alors égal à :

$$p_n = \frac{\frac{\varsigma}{e_m} - 1}{\sqrt{\frac{n}{2} - 1}} = \frac{\sqrt{\frac{n \sum e^2}{4\,\alpha^2}}}{\sqrt{\frac{n}{2} - 1}}\,. \qquad [76]$$

Pour les séries périodiques monocéphales à 12, 24 et 7 termes, le coefficient de précipitation sera égal (pour les détails du calcul, voyez § 48)

$$p_{12} = 0{,}69 \left(\frac{\varsigma}{e_m} - 1\right) = 0{,}69 \left(\sqrt{\frac{3 \sum e^2}{\alpha^2}} - 1\right) \qquad [77]$$

$$p_{24} = 0{,}406 \left(\frac{\varsigma}{e_m} - 1\right) = 0{,}406 \left(\sqrt{\frac{6 \sum e^2}{\alpha^2}} - 1\right) \qquad [78]$$

$$p_7 = 1{,}148 \left(\frac{\varsigma}{e_m} - 1\right) = 1{,}148 \left(\sqrt{\frac{7 \sum e^2}{4\,\alpha^2}} - 1\right). \qquad [79]$$

Les tables I, *A* et I, *B* de l'Annexe I, qui donnent directement c_n d'après $\frac{\sigma}{d_m}$ ou d'après $\frac{\sum d^2}{\sum^2 d}$, donnent en même temps aussi les coefficients de la précipitation des séries périodiques p_n d'après $\frac{\varsigma}{e_m}$ ou d'après $\frac{\sum e^2}{\alpha^2}$ pour des séries à 12, à 24 et à 7 termes.

§ 56. — Comme pour les écarts moyennaux, au lieu de calculer les indices ou les coefficients de l'inégalité des écarts (concentration là et précipitation ici), nous pouvons calculer les indices et coefficients du degré de l'égalité des écarts (nivellement là et gradualité ici). Pour

les indices et le coefficient de la gradualité des variations des séries monocéphales périodiques le procédé est parfaitement analogue à celui employé au § 49 pour le nivellement. Nous pouvons ainsi écrire d'après [74] :

$$1 \geqq \frac{e_m}{\varsigma} \geqq \sqrt{\frac{2}{n}} ; \qquad [80]$$

d'où l'on trouve :

$$1 \geqq \frac{\frac{e_m}{\varsigma} - \sqrt{\frac{2}{n}}}{1 - \sqrt{\frac{2}{n}}} \geqq 0 . \qquad [81]$$

Le *coefficient de la gradualité* des variations d'une série périodique monocéphale, que nous désignerons par g, s'exprime alors par la formule :

$$g_n = \frac{\frac{e_m}{\varsigma} - \sqrt{\frac{2}{n}}}{1 - \sqrt{\frac{2}{n}}} \qquad [82]$$

Appliqué aux séries périodiques monocéphales de 12, 24 et 7 termes, le coefficient de gradualité devient :

$$g_{12} = 1{,}69 \left(\frac{e_m}{\varsigma} - 0{,}408\right) = 1{,}69 \left(\sqrt{\frac{\alpha^2}{3 \Sigma e^2}} - 0{,}408\right) \qquad [83]$$

$$g_{24} = 1{,}41 \left(\frac{e_m}{\varsigma} - 0{,}289\right) = 1{,}41 \left(\sqrt{\frac{\alpha^2}{6 \Sigma e^2}} - 0{,}289\right) \qquad [84]$$

$$g_7 = 2{,}15 \left(\frac{e_m}{\varsigma} - 0{,}535\right) = 2{,}15 \left(\sqrt{\frac{4 \alpha^2}{7 \Sigma e^2}} - 0{,}535\right). \qquad [85]$$

Notons encore que les observations faites plus haut (§§ 50 à 53) concernant la concentration et le nivellement des séries s'appliquent également et pour les mêmes raisons, à la précipitation et à la gradualité des séries périodiques monocéphales. Ici comme plus haut, les indices inachevés ne doivent être appliqués qu'aux séries homonomes ; les coefficients de la précipitation et de la gradualité (comme ceux de

la concentration et du nivellement) ne sont pas complémentaires et leur somme est inférieure à 1 (sauf dans les deux cas limites) ; la comparaison des coefficients de précipitation et de gradualité de séries hétéronomes n'est légitime que s'il s'agit de séries *objectivement* hétéronomes ; des séries hétéronomes ayant le même indice $\frac{\varsigma}{e_m}$ $\left(\text{ou } \frac{e_m}{\varsigma}\right)$ diffèrent moins par leur coefficient de gradation que par leur coefficient de précipitation, et cela dans la mesure indiquée par la formule [69].

Ajoutons enfin que toutes les formules déduites ici pour les séries monocéphales périodiques s'appliquent aussi aux séries monocéphales *non périodiques* avec les deux corrections suivantes : 1° n signifie partout le nombre des écarts successifs qui, pour des séries non périodiques, est égal au *nombre des termes moins* 1, et 2° les α de nos formules doivent partout être remplacés par $\frac{\Sigma e}{2}$ ou (selon § 35, note 2) par $\alpha - \frac{[v_n - v_1]}{2}$.

§ 57. — Nous avons vu (§ 37) qu'on peut établir un indice de concentration également sur la base de l'écartement moyennal du *second* ordre $\left(K = \frac{\delta_m}{d_m} = \frac{\Sigma \delta}{\Sigma d}\right)$. Cet indice croît et décroît directement avec les écarts moyennaux du second ordre, c'est-à-dire avec l'inégalité des écarts moyennaux du premier ordre, soit avec la concentration des variations dans la série. Il atteint son *minimum*, qui est $= 0$, lorsque tous les δ sont $= 0$, c'est-à-dire quand les écarts moyennaux du premier ordre sont tous égaux entre eux, le nivellement de la série étant ainsi parfait et la concentration nulle. Quelle est la valeur *maxima* de cet indice croissant avec la croissance des écarts du second ordre ? Ou, en d'autres termes, que devient l'indice K lorsque la concentration est maxima ?

Comme nous l'avons déjà remarqué (§ 45), la concentration des variations est maxima dans le cas où la série n'accuse qu'un seul écart moyennal positif et un seul négatif, tous les autres écarts étant $= 0$; les deux écarts valables sont alors égaux chacun à la demi-somme de tous les écarts moyennaux (du premier ordre), soit $= \frac{\Sigma d}{2}$. L'écart moyennal du second ordre de chacun des deux écarts valables (pris en valeur absolue) est alors $= \left(\frac{\Sigma d}{2} - d_m\right)$. L'écart moyennal du se-

cond ordre de chacun des $(n - 2)$ écarts restants (égaux à 0) est égal, en valeur absolue, à $(d_m - 0) = d_m$. La somme de tous les écarts du second ordre en cas de concentration maxima est donc égale à :

$$\sum_{\max} \delta = 2\left(\frac{\Sigma d}{2} - d_m\right) + (n-2)\, d_m = \Sigma d - 2\, d_m + n\, d_m - 2\, d_m =$$
$$= \Sigma d + n\, d_m - 4\, d_m\ ;$$

mais $n\, d_m = \Sigma d$; nous pouvons donc écrire :

$$\sum_{\max} \delta = 2\, \Sigma d - 4\, d_m = 2\, (\Sigma d - 2\, d_m) \qquad [86]$$

La limite minima de $\Sigma \delta$ étant 0, nous avons ainsi les deux limites de $\Sigma \delta$; elles sont données par l'expression suivante :

$$(2\, \Sigma d - 4\, d_m) \geqq \Sigma \delta \geqq 0 \qquad [87]$$

L'écart moyennal arithmétique moyen du second ordre (δ_m) varie alors dans les limites que voici :

$$\frac{2\, \Sigma d - 4\, d_m}{n} \geqq d_m \geqq 0$$

ou, ce qui revient au même $\left(\frac{\Sigma d}{n} \text{ étant } = d_m\right)$:

$$2\, d_m \left(1 - \frac{2}{n}\right) \geqq \delta_m \geqq 0\,. \qquad [88]$$

L'indice de concentration $K \left(= \frac{\delta_m}{d_m}\right)$ varie donc, selon la concentration des séries, dans les limites suivantes :

$$2\left(1 - \frac{2}{n}\right) \geqq \frac{\delta_m}{d_m} \geqq 0 \qquad [89]$$

ou, si l'on préfère :

$$\left(2 - \frac{4}{n}\right) \geqq K \geqq 0\,. \qquad [89^{bis}]$$

§ 58. — Connaissant ainsi les limites des variations de K, nous pouvons maintenant, sur la base de l'écartement moyennal du second

ordre, établir également un *coefficient de concentration des écarts*, que nous désignerons par k_n, qui sous la forme d'une fraction positive variant entre 0 et 1, traduit les degrés de concentration possibles en prenant la concentration de toute série par rapport au maximum de concentration possible pour une série ayant le même nombre de termes.

Il résulte, en effet, de la formule [89^bis] que :

$$1 \overline{>} \frac{n}{2n-4} K \overline{>} 0 \,;$$

le coefficient cherché est ainsi égal à :

$$k_n = \frac{n}{2n-4} K = \left(\frac{n}{2n-4}\right) \cdot \frac{\delta_m}{d_m}. \qquad [90]$$

La dernière expression pourrait encore être modifiée par la substitution de $\Sigma\,\delta$ à $(n\,\delta_m)$; on aurait alors :

$$k_n = \frac{\Sigma\,\delta}{(2n-4)\,d_m} = \frac{\Sigma\,\delta}{2\,\Sigma\,d - 4\,d_m} = \frac{n}{2\,(n-2)} \cdot \frac{\Sigma\,\delta}{\Sigma\,d}. \qquad [91]$$

Le passage de l'indice de concentration K au coefficient k_n est, comme on le voit d'après la formule [90], excessivement simple : on n'a qu'à multiplier celui-là par $\frac{n}{2n-4}$ pour obtenir celui-ci (n étant toujours le nombre des termes). On aura ainsi pour les séries à 12, à 24 et à 7 termes, les formules suivantes de k_n :

$$k_{12} = \frac{12}{20} \cdot \frac{\delta_m}{d_m} = 0{,}6\,\frac{\delta_m}{d_m} \qquad [92]$$

$$k_{24} = \frac{24}{44} \cdot \frac{\delta_m}{d_m} = \frac{6}{11} \cdot \frac{\delta_m}{d_m} = 0{,}55\,\frac{\delta_m}{d_m} \qquad [93]$$

$$k_7 = \frac{7}{10} \cdot \frac{\delta_m}{d_m} = 0{,}7\,\frac{\delta_m}{d_m}. \qquad [94]$$

Ajoutons encore que le coefficient k_n, déduit sans l'introduction de considérations propres aux seules séries périodiques, vaut pour les séries ordinaires comme pour les séries périodiques.

§ 59. — Ici encore nous voyons, d'après [89] ou [89 bis], que le maximum de l'*indice* de concentration possible varie avec le nombre des termes de la série $\left(K \overline{<} 2 - \frac{4}{n}\right)$. Le même indice de concentration K ne signifie donc pas un degré de concentration égal pour des séries hétéronomes. Nous comprenons maintenant pourquoi l'indice K (comme les autres indices de la concentration) ne nous donne une idée exacte de la concentration plus ou moins forte des séries que lorsqu'il s'agit de comparer des séries homonomes. Quand il s'agit de séries hétéronomes, c'est toujours le coefficient (k_n) qu'il faut donc envisager.

Cependant, en appliquant l'indice K à des séries hétéronomes nous commettons une erreur moindre qu'en appliquant l'un des indices basés sur les carrés des écarts du premier ordre. En effet, sous l'influence du nombre des termes, le maximum de K varie beaucoup moins que celui des autres indices de concentration. Ainsi, tandis que le maximum de $\frac{\sigma}{d_m}$ est d'après [50] $= \sqrt{\frac{n}{2}}$, devenant ainsi indéfiniment grand avec l'augmentation de n, tandis que le maximum de C_n est même, d'après [53] $= \frac{n}{2}$, le maximum de K ne peut, d'après [89 bis], jamais dépasser le chiffre de 2 (il ne devient $= 2$ que si $n = \infty$). Aussi, la réduction que le coefficient k_n apporte à l'indice K [90] ne va-t-elle pas au-delà de $\frac{1}{2}$ $\left(\text{En effet, } \frac{n}{2n-4} = \frac{1}{2 - \frac{4}{n}}\right)$. Nous avons vu, d'après [92] et [93], que pour les séries les plus usitées, celles de 12 et de 24 termes, les coefficient k_n est égal respectivement à 0,60 K et à 0,55 K ; pour une série à 48 termes, $k_{48} = 0{,}52\ K$, et pour aucune série finie, quelque grand que soit le nombre de ses termes, la correction de K ne tomberait à 0,50. On peut donc dire que pour des séries ayant plus de 24 termes, les indices K sont pratiquement comparables, même s'il s'agit de séries hétéronomes. Avec la diminution du nombre des termes, la correction nécessaire de K, pour la comparaison de séries hétéronomes, augmente. Mais ici encore nous avons vu [94] que $k_7 = 0{,}7\ K$; la différence de correction n'est donc pas excessivement grande entre les séries à 7 termes et les séries à 12 termes ($k_{12} = 0{,}6\ K$) ; elle est en tout cas moindre que pour les indices de concentration basés sur les carrés des écarts du premier ordre.

§ 60. — Envisageons maintenant de plus près la nature des variations du coefficient k_n en fonction de la concentration des séries. Examinons notamment les variations de k_n pour le cas des séries ayant le même nombre de termes et la même somme d'écarts moyennaux et, de plus, dans l'hypothèse qu'aucun des écarts moyennaux du premier ordre qui ne sont pas égaux à 0, n'est inférieur (en valeur absolue) à l'écart moyennal arithmétique moyen. Autrement dit, admettons n et Σd constants et d soit $\geq d_m$ soit $= 0$; voyons quelles seraient dans ces conditions les variations de k_n.

D'après [91], nous avons : $k_n = \frac{n}{2\,(n - 2)} \cdot \frac{\Sigma\,\delta}{\Sigma\,d}$; dans cette équation, seule l'expression $\Sigma\,\delta$ varie avec la concentration de la série ; examinons donc $\Sigma\,\delta$. Admettons que toute la somme des écarts moyennaux du premier ordre ($\Sigma\,d$) se trouve concentré sur un nombre de termes l et que tous les autres termes de la série (selon notre hypothèse) ne s'écartent en rien de la moyenne de la série (v_m), c'est-à-dire que les $(n - l)$ écarts moyennaux qui restent sont tous $= 0$. Que devient alors $\Sigma\,\delta$?

L'écart du second ordre de chaque terme (δ) étant égal en valeur absolue à la différence entre l'écart moyennal du premier ordre du même terme et l'écart moyennal arithmétique moyen $[d - d_m]$, la somme des écarts du second ordre des l termes sur lesquels se trouve concentrée toute la somme des écarts $\Sigma\,d$ est égale à la somme de leurs écarts du premier ordre moins l fois l'écart moyennal arithmétique moyen ; en désignant la somme des écarts du second ordre de ces l termes par $\Sigma'\,\delta$, nous pouvons donc écrire :

$$\Sigma'\,\delta = \Sigma\,d - l\,d_m \qquad (1)$$

Les $(n - l)$ termes qui restent accusant chacun $d = 0$, ont chacun un écart du second ordre égal en valeur absolue à $[0 - d_m =]\ d_m$; la somme des écarts du second ordre de ces $(n - l)$ termes ($\Sigma''\,\delta$) est ainsi égale à :

$$\Sigma''\,\delta = (n - l)\ d_m = n\,d_m - l\,d_m = \Sigma\,d - l\,d_m \qquad (2)$$

En additionnant les équations (1) et (2), nous obtenons donc :

$$\Sigma\,\delta\ (= \Sigma'\,\delta + \Sigma''\,\delta) = 2\,(\Sigma\,d - l\,d_m) = 2\left(\Sigma\,d - l\,\frac{\Sigma\,d}{n}\right) =$$

$$= \frac{2}{n}\,(n - l)\ \Sigma\,d\,. \qquad [95]$$

Substituant, dans la formule du coefficient k_n [91], à $\Sigma\delta$ sa valeur d'après [95], nous trouvons :

$$k_n = \frac{n}{2(n-2)} \cdot \frac{\Sigma\delta}{\Sigma d} = \frac{n \,.\, 2(n-l)\,\Sigma d}{2(n-2)\,\Sigma d \,.\, n} = \frac{n-l}{n-2}. \qquad [96]$$

n étant constant, la formule [96] nous montre ainsi les variations de k_n sous la seule influence des variations du nombre des termes (l) sur lesquels se trouve concentrée toute la somme des écarts moyennaux. Cette formule peut d'ailleurs s'écrire encore sous la forme suivante :

$$k_n = \frac{n}{n-2} - \frac{l}{n-2}. \qquad [96^{bis}]$$

Dans notre hypothèse (n constant et les d inégaux à 0 étant $\overline{>} d_m$) cela veut dire que si le nombre des termes sur lequel l'écartement moyennal est concentré (l) est de 1 plus grand, le coefficient k_n est de $\frac{1}{n-2}$ plus petit, s'il est plus grand de 2, le coefficient est plus petit de $\frac{2}{n-2}$, et ainsi de suite ; autrement dit : *pour les séries ayant le même nombre de termes et dans l'hypothèse que tous les* d *inégaux à* 0 *sont* $\overline{>} d_m$, *les variations du coefficient de concentration* k_n *sont inversement proportionnelles aux variations du nombre des termes sur lesquels l'écartement moyennal est concentré.*

Nous pouvons dire encore que dans l'hypothèse indiquée, k_n varie *en ligne droite* et en sens inverse du nombre des termes sur lesquels est concentré l'écartement moyennal.

Ainsi (toujours dans la même hypothèse), si n est par exemple $=$ 12, k_n diminue de 0,1 $\left(= \frac{1}{12-2}\right)$ pour toute extension de l'écartement moyennal sur 1 terme de plus. Le nombre minimum de termes sur lequel s'étend l'écartement moyennal (cas de concentration maxima) étant $=$ 2 (soit un écart positif et un négatif), nous avons de la sorte d'après [96] :

pour $l = 2, \ldots\ k_{12} = \frac{12-2}{12-2} = 1$

» $l = 3, \ldots\ k_{12} = \frac{12-3}{12-2} = 0{,}90$

$$\text{pour } l = 4, \ldots\ k_{12} = \frac{12-4}{10} = 0{,}80, \text{ et ainsi de suite ;}$$

$$\text{» } l = 11, \ldots\ k_{12} = \frac{12-11}{10} = 0{,}10$$

$$\text{» } l = 12, \ldots\ k_{12} = \frac{12-12}{10} = 0 \text{ (cas de concentration nulle).}$$

Cette simplicité de la loi des variations de k_n, cette variation inversement *proportionnelle* aux variations du nombre des termes sur lesquels la somme des écarts moyennaux est concentrée, cette variation en ligne droite confèrerait certainement au coefficient de concentration k_n une énorme supériorité sur le coefficient c_n, dont la loi de variation paraît beaucoup moins simple.

§ 61. — Cependant, la formule [96] nous montre encore autre chose. Il résulte en effet de cette formule que, dans l'hypothèse que tous les d inégaux à 0 sont $\overline{>} d_m$, le coefficient k_n (n étant constant) dépend *uniquement* de l, c'est-à-dire du *nombre* des termes sur lesquels l'écartement moyennal est concentré ; *toujours dans la même hypothèse, k_n est donc insensible à l'égalité ou à l'inégalité des écarts moyennaux des* l *termes sur lesquels l'écartement moyennal est concentré ;* autrement dit, *dans l'hypothèse indiquée*, k_n se révèle insensible à certaines formes importantes de la concentration des séries, à l'intensité de la concentration parmi les termes accusant un écart moyennal inégal à 0. — Il n'en est pas de même du coefficient c_{12} qui, basé sur σ, sur les *carrés* des écarts moyennaux (du premier ordre), se montre sensible à toute inégalité dans la répartition de la somme des écarts moyennaux de la série, c'est-à-dire à toute variation de la concentration.

Pour rendre cette différence entre k_n et c_n plus tangible, voici quelques exemples répondant à l'hypothèse posée. Prenons notamment les séries suivantes à 12 termes ($n = 12$). (Rappelons que v = termes de la série, d = écarts moyennaux du premier ordre, δ = écarts moyennaux du second ordre) :

a) $l = 4$; $\Sigma d = 24$.

Série I.

													En moyenne
v	14,	14,	26,	26,	20,	20,	20,	20,	20,	20,	20,	20	20
d	− 6,	− 6,	+ 6,	+ 6,	0,	0,	0,	0,	0,	0,	0,	0	2
δ	+ 4,	+ 4,	+ 4,	+ 4,	− 2,	− 2,	− 2,	− 2,	− 2,	− 2,	− 2,	− 2	$\frac{32}{12} = \frac{8}{3}$

Série II.

													En moyenne
v	8,	28,	22,	22,	20,	20,	20,	20,	20,	20,	20,	20	20
d	-12,	$+8$,	$+2$,	$+2$,	0,	0,	0,	0,	0,	0,	0,	0	2
δ	$+10$,	$+6$,	0,	0,	-2,	-2,	-2,	-2;	-2,	-2,	-2,	-2	$\frac{32}{12} = \frac{8}{3}$

Pour le coefficient k_{12}, on trouve d'après [92] dans les deux cas :

$$k_{12} = 0{,}6 \frac{\delta_m}{d_m} = \frac{0{,}6 \,.\, \frac{8}{3}}{2} = 0{,}80 \,.$$

On obtiendrait naturellement le même résultat d'après [96] :

$$k_{12} = \frac{12 - 4}{12 - 2} = 0{,}80 \,.$$

Par contre, le coefficient c_{12} est d'après [59] :

pour la série I :

$$c_{12} = 0{,}69 \left(\frac{\sigma}{d_m} - 1\right) = 0{,}69 \left(\frac{\sqrt{\frac{36 \,.\, 4}{12}}}{2} - 1\right) = 0{,}69\,(\sqrt{3} - 1) =$$

$$= 0{,}69 \,.\, 0{,}732 = 0{,}505$$

pour la série II :

$$c_{12} = 0{,}69 \left(\frac{\sigma}{d_m} - 1\right) = 0{,}69 \left(\frac{\sqrt{\frac{144 + 64 + 4 + 4}{12}}}{2} - 1\right) =$$

$$= 0{,}69 \left(\sqrt{\frac{216}{12 \,.\, 4}} - 1\right) = 0{,}69\,(\sqrt{4{,}5} - 1) = 0{,}69 . 1{,}1213 = 0{,}774 \,.$$

Or, il est évident que la concentration de l'écartement moyennal est loin d'être la même dans les deux séries. Dans la série II, les $\frac{5}{6}$ de toute la somme des écarts moyennaux $\left(\frac{12 + 8}{24}\right)$ se trouvent con-

centrés sur deux termes de la série, tandis que dans la série I, sur les deux termes accusant les plus forts écarts moyennaux se trouve concentrée seulement la moitié de toute la somme des écarts $\left(\frac{6+6}{24}\right)$; même avec trois termes on n'arrive pas encore à la proportion de $\frac{5}{6}$ $\left(\text{en effet, } \frac{6+6+6}{24} = \frac{3}{4}\right)$. Cette grande différence de concentration apparaît avec le coefficient c_{12}, mais se trouve entièrement voilée avec k_{12}.

Le résultat serait naturellement le même si l'on prenait des séries ayant une autre somme d'écarts moyennaux (puisque notre formule [96] est indépendante de $\Sigma\, d$).

Prenons d'ailleurs encore deux autres séries, différant des deux séries que nous venons d'examiner par la valeur de l et de Σd (nous prenons toujours des séries à 12 termes, mais il est évident qu'on pourrait prendre tout aussi bien des séries ayant n'importe quel autre nombre de termes).

b) $l = 8$; $\Sigma d = 36$.

Série III.

													En moyenne
v	15,	16,	16,	15,	20,	20,	24,	25,	25,	24,	20,	20	20
d	− 5,	− 4,	− 4,	− 5,	0,	0,	+ 4,	+ 5,	+ 5,	+ 4,	0,	0	3
δ	+ 2,	+ 1,	+ 1,	+ 2,	− 3,	− 3,	+ 1,	+ 2,	+ 2,	+ 1,	− 3,	− 3	2

Série IV :

v	17,	11,	17,	17,	20,	20,	23,	24,	28,	23,	20,	20	20
d	− 3,	− 9,	− 3,	− 3,	0,	0,	+ 3,	+ 4,	+ 8,	+ 3,	0,	0	3
δ	0,	+ 6,	0,	0,	− 3,	− 3,	0,	+ 1,	+ 5,	0,	− 3,	− 3	2

Le coefficient k_{12} donne dans les deux cas :

$$k_{12} = 0{,}6\,\frac{\delta_m}{d_m} = \frac{0{,}6\,.\,2}{3} = 0{,}40\,.$$

(De même d'après [96], on aurait dans les deux cas :

$$k_{12} = \frac{n-l}{n-2} = \frac{12-8}{10} = 0{,}40\Big)\,.$$

Par contre, c_{12} devient :

pour la série III :

$$c_{12} = 0{,}69 \left(\frac{\sigma}{d_m} - 1\right) = \left(\frac{\sqrt{\dfrac{25 \cdot 4 + 16 \cdot 4}{12}}}{3} - 1\right) . \, 0{,}69 =$$

$$= 0{,}69 \, . \left(\sqrt{\frac{41}{27}} - 1\right) = 0{,}69 \left(\sqrt{1{,}5185} - 1\right) = 0{,}69 \, . \, 0{,}23 = 0{,}159$$

pour la série IV :

$$c_{12} = 0{,}69 \left(\frac{\sigma}{d_m} - 1\right) = 0{,}69 \left(\frac{\sqrt{\dfrac{9 \cdot 5 + 81 + 16 + 64}{12}}}{3} - 1\right) =$$

$$= 0{,}69 \left(\sqrt{\frac{206}{108}} - 1\right) = 0{,}69 \left(\sqrt{1{,}9074} - 1\right) = 0{,}69 \, . \, 0{,}38 = 0{,}262 \, .$$

Ici encore, les deux séries ont une concentration fort différente. Dans la série IV, près de la moitié de la somme des écarts moyennaux $\left(\frac{9+8}{36} = \frac{17}{36}\right)$ se trouve concentrée sur *deux* termes tandis que dans la série III, même en prenant *trois* termes accusant les écarts maxima nous n'obtenons que $\frac{15}{36}$ de la somme de tous les écarts. Et pourtant k_{12} donne dans les deux cas le même chiffre : 0,40 ; par contre, le coefficient c_{12} donne 0,26 pour la série IV et seulement 0,16 pour la série III.

En présence de ces constatations, il nous semble établi que le *véritable coefficient de la concentration des écarts est bien c_n, basé sur les carrés des écarts moyennaux du premier ordre ; il se montre plus juste que k_n, basé sur les écarts moyennaux du second ordre.*

Nous ne croyons pourtant pas que k_n doive être purement et simplement rejeté. Dans la réalité, il traduit encore assez bien les différences de concentration de différentes séries. Il a en outre des qualités incontestables de clarté, de simplicité et de facilité de calcul. Seulement, quand il y a divergence au sujet de la plus ou moins grande concentration des séries comparées selon que l'on applique le coefficient c_n ou k_n, c'est l'indication fournie par c_n qui doit être considérée comme exacte.

On pourrait peut-être désigner k_n comme le coefficient *simple* de la concentration et c_n comme le coefficient de concentration *qualifié* (on serait presque tenté de dire *pondéré*, tant l'analogie est sensible entre k_n et c_n, d'un côté, et les nombres indices simples et pondérés, de l'autre ; nous évitons cependant ici le terme de « pondéré » pour ne pas créer de confusion). En tout cas, si l'on parle du coefficient de concentration sans le spécifier davantage il faudra comprendre par ce terme le coefficient c_n.

§ 62. — Notons enfin encore une différence entre les coefficients c_n et k_n. Le lecteur a probablement remarqué que dans les exemples cités, k_n s'exprime continuellement par un chiffre plus élevé que c_n (pour la série I, $k_{12} = 0{,}80$ et $c_{12} = 0{,}51$; pour la série II, $k_{12} = 0{,}80$ et $c_{12} = 0{,}77$; pour III, $k_{12} = 0{,}40$ et $c_{12} = 0{,}16$; pour IV, $k_{12} = 0{,}40$ et $c_{12} = 0{,}26$). Ceci n'est pas un hasard. Nous avons déjà remarqué (§ 37) que l'indice de concentration basé sur les écarts du second ordre (K) croît pour les petites concentrations beaucoup plus fortement que les indices basés sur les carrés des écarts du premier ordre $\left(\frac{\sigma}{d_m}, C_n\right)$, et plus faiblement que ceux-ci pour les grandes concentrations. Il en est de même des coefficients k_n et c_n qui sont fonctions positives et directes de ces indices. Pour les grandes concentrations, la différence entre k_n et c_n s'atténue, et quand le concentration est maxima les deux deviennent $= 1$. Dans l'hypothèse que nous venons d'examiner dans les deux derniers §§, k_n présente une ligne droite ; c_n prend alors l'allure d'une courbe concave qui rejoint cette droite à ses deux points extrêmes (0 et 1), mais qui se trouve au-dessous d'elle sur tout le reste de son étendue. — Le fait que k_n se montre bien plus sensible que c_n aux petites concentrations (qui dans la réalité sont très fréquentes) est peut-être la raison la plus importante pour ne pas abandonner complètement k_n comme mesure de concentration (*).

Il va de soi que k_n, tout comme c_n, est applicable aussi bien aux séries périodiques qu'aux non-périodiques.

(*) Dans une étude récente sur *les Fluctuations saisonnières dans l'industrie du bâtiment* (« Revue trimestrielle de statistique » publiée par l'Office Central de Statistique de la Rép. Polonaise, 1933, fasc. 2-3), le Dr. J. WISNIEWSKI, tout en faisant un large usage de notre propre étude sur le même sujet mentionnée plus haut, s'est cependant refusé apppliquer notre cœfficient de la concentration des écarts C_{12} (que nous avions d'abord désigné par β) parce que pratiquement,

CHAPITRE IX.

Coefficients de certains caractères des séries périodiques (*suite*).

§ 63. — Comme nous l'avons remarqué au début du précédent chapitre, notre *indice de dissymétrie D* se rapproche d'un coefficient dans ce sens que, variant entre 0 et 1, il marque le *degré* atteint par le phénomène par rapport au maximum qu'il peut en général atteindre. Il ne nous paraît cependant pas encore de nature à nous satisfaire entièrement.

En effet, dans notre conception de la symétrie et de la dissymétrie, toutes les séries dont les termes équidistants du centre de symétrie ne varient pas toujours dans le même sens ni toujours dans un sens opposé, offrent à la fois une certaine déviation de la symétrie positive et une certaine autre de la symétrie négative (dont l'indice est > 0 et < 1). En principe il faudrait donc calculer toujours les deux indices de dissymétrie, le positif et le négatif (sans parler des trois genres de symétrie possibles, symétrie des écarts moyennaux, successifs et centraux, qui ne sont pas des notions opposées, mais qui correspondent à des aspects différents de la réalité). Comme nous avons vu, dans bien des cas, notamment dans les cas où la prédominance de la symétrie positive ou de la symétrie négative n'est pas certaine d'emblée, il faut effectivement calculer les deux indices (§ 41, séries III et IV). Mais ce qui est plus important encore, c'est qu'à un même indice de dissymétrie d'un sens donné (positive ou négative), peuvent correspondre différents indices de dissymétrie en sens opposé. Ceci est évident quand l'indice de dissymétrie est *égal à* 1 ; car, comme nous l'avons vu au § 40, un indice égal à 1 peut être obtenu dans deux cas : 1° lorsque dans tous les couples de termes équidistants l'un des termes reste invariable tandis que l'autre varie, et 2° lorsque dans tous les

---

notamment pour le phénomène étudié, ce cœfficient prend des valeurs peu élevées. Cette raison, il est vrai, ne nous paraît pas péremptoire ; mais dans la pratique, pour certains phénomènes, le cofficient k_n conviendrait peut-être mieux. Voyez aussi plus loin, chapitre X et notamment § 71 (6°) et § 76.

couples de termes équidistants, l'un des termes varie en sens opposé aux variations de l'autre (s'il s'agit d'un indice de dissymétrie positive) ou s'ils varient dans le même sens (s'il s'agit d'un indice de dissymétrie négative) ; dans le premier cas, l'indice de dissymétrie opposée est lui aussi égal à 1 ; dans le second, il est inférieur à 1 et peut même devenir 0 si les écarts des termes équidistants sont de valeur absolue égale. Il en est de même si l'indice de dissymétrie donné est *inférieur à* 1: l'indice de la dissymétrie opposée peut prendre des valeurs différentes selon que la dissymétrie donnée provient plus ou moins d'une variation *inégale* des termes équidistants ou d'une variation de certains termes équidistants dans un *sens opposé* à celui de la symétrie considérée (car la variation des termes équidistants dans un sens opposé à la symétrie considérée influe aussi sur l'indice de la dissymétrie opposée tandis que la variation inégale de ces termes dans le même sens que la dissymétrie considérée laisse l'indice de la dissymétrie opposée sans changement). Pour nous en rendre compte plus clairement, prenons quelques exemples. Soit quatre séries accusant les écarts (centraux, successifs ou moyennaux) suivants :

I	$+5, -1, -3, +2, 0, +3, -1, -2, +3$
II	$+5, -1, -3, +2, 0, +3, -2, +1, +3$
III	$+4, -2, -3, +2, 0, -2, -3, -1, +3$
IV	$+3, -2, -3, +2, 0, +2, +3, -2, +3$

Les indices de dissymétrie positive (D^+) et négative (D^-) de ces séries sont les suivants :

Série I (tous les écarts des termes équidistants varient dans le même sens) :

$$D_1^+ = \frac{1+2+1+2}{20} = \frac{6}{20} = 0{,}30 \; ; \; D_1^- = \frac{5+4+3+8}{20} = \frac{20}{20} = 1 \, .$$

Série II (deux écarts de termes équidistants varient en sens opposés, l'un étant $= -1$ et l'autre $= +1$) :

$$D_2^+ = \frac{1+1+2+2}{20} = \frac{6}{20} = 0{,}30 \; ; \; D_2^- = \frac{5+5+8}{20} = \frac{18}{20} = 0{,}90 \, .$$

Série III (un écart est $= + 2$, et l'écart du terme équidistant $= - 2$) :

$$D_3^+ = \frac{4+1+1}{20} = \frac{6}{20} = 0{,}30\ ;\ D_3^- = \frac{6+3+7}{20} = \frac{16}{20} = 0{,}80\ .$$

Série IV (un écart $= - 3$, et l'écart du terme équidistant $= + 3$) :

$$D_4^+ = \frac{6}{20} = 0{,}30\ ;\ D_4^- = \frac{4+4+6}{20} = \frac{14}{20} = 0{,}70\ ;$$

et ainsi de suite.

Il faut donc, à proprement parler, marquer les deux indices même quand la prédominance de la symétrie positive ou de la négative est de toute évidence comme dans les exemples que nous venons d'examiner. On pourrait le faire aisément en mettant entre parenthèses l'indice de dissymétrie le plus élevé (qui marque la symétrie dont la série s'écarte le plus) en le faisant précéder du signe de la symétrie dont il marque la déviation ; il devient alors inutile de marquer d'un signe (+ ou —) la lettre D. Ainsi pour nos séries, on aurait :

I . . . . $D_1 = 0{,}30\ (-1)$
II . . . . $D_2 = 0{,}30\ (-0{,}90)$
III . . . . $D_3 = 0{,}30\ (-0{,}80)$
IV . . . . $D_4 = 0{,}30\ (-0{,}70)$, et ainsi de suite.

Si une série a une prédominance de symétrie négative, si par exemple ses indices de dissymétrie sont respectivement : $D^+ = 0{,}90$ et $D^- = 0{,}20$, on écrirait :

$$D = -0{,}20\ (+0{,}90)\,.$$

Comme le calcul de ces indices est on ne peut plus simple, l'emploi d'un indice double ne présenterait en somme aucune difficulté.

Cependant, l'image serait moins nette qu'avec un chiffre unique qui résumerait le tout.

Cette image paraîtrait insuffisamment nette surtout si, en comparant deux séries, nous trouvions pour elles des indices qui tous les deux seraient pour l'une des séries plus élevés que pour l'autre, par exemple :

$$D_1 = 0{,}30\ (-0{,}80)$$
$$D_2 = 0{,}25\ (-0{,}75).$$

En réalité, cela signifierait que les deux séries s'écartent relativement peu de la symétrie positive, que la seconde s'en écarte moins que la première, mais que cette déviation est dans la seconde plus que dans la première l'effet d'une variation des termes équidistants dans un sens opposé. On ne peut pas nier que ces détails de la dissymétrie aient quelque utilité eux aussi. Mais ce qu'on gagne ici en précision du détail, le résumé d'ensemble le perd en netteté.

§ 64. — Au lieu de fixer le degré de la *dissymétrie* des séries, nous pouvons chercher à déterminer directement le degré de leur *symétrie*. Quant au degré de concordance (ou de discordance) des variations des termes équidistants, au lieu de le déterminer en prenant pour base la différence (ou la somme algébrique) de leurs écarts par rapport à la somme de ces mêmes écarts, on peut l'établir en se basant sur *les produits de ces écarts comparés aux moyennes de leurs carrés.*

En effet, désignons les écarts (moyennaux, successifs ou centraux) des termes situés à gauche du centre de symétrie par x et ceux des termes situés à droite par y (x et y pouvant être positifs ou négatifs) et prenons les écarts d'*un* couple de termes équidistants, par exemple x_n et y_n. Si ces termes accusent une parfaite symétrie positive, alors $x_n = y_n$ et $x_n\, y_n = \dfrac{x_n^2 + y_n^2}{2}$ (puisque chaque membre de cette expression est $= x_n^2 = y_n^2$) ; dans ce cas, le produit des écarts des deux termes équidistants est donc égal à la demi-somme des carrés de ces écarts ; autrement dit, le rapport $\dfrac{x_n\, y_n}{(x_n^2 + y_n^2) : 2} = 1$. Pour une somme $(x + y)$ donnée $(= S)$, et quelle que soit sa valeur, le produit $x\,y$ est *maximum* quand $x = y$ (d'après § 44, proposition 3) et il est alors égal à $\left(\dfrac{S}{2}\right)^2$; pour la même somme (S), si $x = y$, la somme des carrés $(x^2 + y^2)$ est alors *minima* (proposition 2 du même § 44) et elle est alors égale à $x^2 + y^2 = \dfrac{S^2}{2}$. Le rapport $\dfrac{x\,y}{x^2 + y^2}$ est donc égal à $\dfrac{S^2}{4} : \dfrac{S^2}{2} = \dfrac{1}{2}$ quand le numérateur $(x\,y)$ est maximum et le dénominateur $(x^2 + y^2)$ est minimum, soit quand la fraction est maxima. Nous pouvons ainsi écrire :

$$\frac{x\,y}{x^2 + y^2} \leqq \frac{1}{2}, \qquad [97]$$

ou encore :

$$\frac{x\,y}{(x^2+y^2):2} \leqq 1 \qquad \text{ou bien :} \qquad x\,y \leqq \frac{x^2+y^2}{2}. \qquad [97^{bis}]$$

La valeur 1 pour le rapport $\frac{x_n\,y_n}{(x_n^2+y_n^2):2}$ qui se vérifie dans le cas de symétrie parfaite d'un couple de termes équidistants ($x = y$) constitue ainsi un maximum. A mesure que l'inégalité entre x_n et y_n augmente (que la symétrie diminue), le numérateur ($x_n\,y_n$) diminue et en même temps le dénominateur ($x_n^2 + y_n^2$) augmente, c'est-à-dire le rapport $\frac{x_n\,y_n}{(x_n^2+y_n^2):2}$ diminue. Enfin, si x_n ou y_n devient $= 0$, c'est-à-dire s'il n'y a aucune symétrie entre x_n et y_n, notre rapport devient lui aussi $= 0$.

Lorsque $x_n = -y_n$, c'est-à-dire en cas de parfaite symétrie négative, on trouve de façon analogue :

que le rapport $\frac{x_n\,y_n}{(x_n^2+y_n^2):2} = -1$;

que -1 est la valeur absolue maxima du rapport considéré ;

qu'avec l'augmentation de l'inégalité des valeurs absolues de x_n et de y_n (qui conservent des signes opposés), le rapport $\frac{x_n\,y_n}{(x_n^2+y_n^2):2}$, tout en demeurant négatif, diminue comme valeur absolue ;

et enfin que si x_n ou y_n devient $= 0$, tout le rapport $\frac{x_n\,y_n}{(x^2+y^2):2}$ devient $= 0$.

Cela veut dire qu'en cas de symétrie négative des écarts x_n et y_n, le rapport $\frac{x_n\,y_n}{(x_n^2+y_n^2):2}$ présente une quantité négative variant entre 0 et -1, étant $= 0$ quand il n'y a point de symétrie entre x_n et y_n, augmentant en valeur absolue avec l'augmentation de la symétrie négative entre eux et devenant $= -1$ quand ils présentent une parfaite symétrie négative.

Ainsi pour *un* couple de termes équidistants, le rapport $\frac{x_n\,y_n}{(x_n^2+y_n^2):2}$ marque tous les degrés de symétrie positive ou négative, variant entre $+1$ et 0 pour la symétrie positive et entre -1 et 0 pour la symétrie négative.

Prenons maintenant *plusieurs* couples de termes équidistants dont les écarts seront pour les termes *gauches* $x_1\ x_2,\ x_3, \ldots x_n$ et ceux des termes droits, $y_1,\ y_2,\ y_3, \ldots y_n$. Voyons quelles seront les variations du rapport :

$$\frac{x_1 y_1 + x_2 y_2 + x_3 y_3 + \ldots + x_n y_n}{\frac{x_1^2 + y_1^2}{2} + \frac{x_2^2 + y_2^2}{2} + \frac{x_3^2 + y_3^2}{2} + \ldots + \frac{x_n^2 + y_n^2}{2}}.$$

Dans le cas où $x_1 = y_1,\ x_2 = y_2,\ x_3 = y_3, \ldots x_n = y_n$, c'est-à-dire dans le cas de parfaite symétrie positive, toute cette expression se transformera en $\frac{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2}{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2} = 1$. Si les x et les y des termes équidistants (ou certains de ces x et y), tout en conservant les mêmes signes, ne sont pas égaux, alors chaque terme ainsi affecté du numérateur est inférieur au terme correspondant du dénominateur et le rapport entier devient par suite une fraction positive < 1 ; à mesure que la différence entre les x et les y (ayant les mêmes signes) des termes équidistants augmente, notre rapport diminue ; il en sera de même si certains x ont des signes opposés à ceux des y des termes équidistants : les produits de ces x et y devenant négatifs réduisent d'autant la somme des produits xy qui forme le numérateur de notre rapport ; autrement dit, à mesure que la symétrie positive de la série diminue, soit par suite de l'augmentation de l'inégalité des écarts des termes équidistants, soit parce que certains de ces écarts prennent des signes opposés, notre rapport diminue. Enfin, si dans tous les couples de termes équidistants l'un des écarts (x ou y) devient zéro tandis que l'autre conserve une certaine valeur, ou encore si dans certains couples les écarts des termes équidistants ont les mêmes signes et dans d'autres ils ont des signes opposés de façon que la somme des produits des uns soit égale en valeur absolue à la somme des produits des autres, le numérateur de notre rapport devient $= 0$; autrement dit, quand il n'y a point de symétrie entre les termes équidistants, soit que dans tous les couples l'un des termes reste invariable tandis que l'autre varie, soit que ces termes n'accusent pas plus de symétrie positive que de symétrie négative, notre rapport devient $= 0$.

De même, si $x_1 = -y_1,\ x_2 = -y_2,\ x_3 = -y_3, \ldots x_n = -y_n$, notre rapport devient :

$\frac{-x_1^2 - x_2^2 - x_3^2, \ldots - x_n^2}{x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2} = -1$. Si les x et les y des termes équidistants

conservant leurs signes opposés, ne sont pas égaux en valeur absolue, notre rapport devient une fraction négative dont la valeur varie entre — 1 et 0 ; sa valeur absolue diminuera avec l'augmentation de l'inégalité des écarts des termes équidistants et avec l'accroissement du nombre et de l'importance des écarts ayant les mêmes signes. Enfin, quand dans tous les couples de termes équidistants, l'un des écarts devient = 0 tandis que l'autre conserve une certaine valeur ou quand les x ont tantôt le même signe que les y correspondants tantôt des signes opposés de façon que la somme des produits des uns soit égale en valeur absolue à la somme des produits des autres, le rapport devient = 0.

En résumé, *notre rapport varie avec la symétrie des séries entre* + 1 *et* 0 *pour tous les degrés de symétrie positive et entre* — 1 *et* 0 *pour tous les degrés de symétrie négative ;* il présente un vrai *coefficient de symétrie ;* nous le désignerons désormais par s. Nous pouvons ainsi écrire :

$$s = \frac{\Sigma\, x\, y}{\dfrac{\Sigma\, x^2 + \Sigma\, y^2}{2}} \qquad [98]$$

§ 65. — Le lecteur aura déjà certainement remarqué la frappante analogie qui existe entre la symétrie telle que nous l'avons exposée et la *corrélation*, entre notre raisonnement sur la mensuration du degré de symétrie et la mensuration du degré de corrélation, entre le coefficient de symétrie s que nous venons de dégager et le coefficient de corrélation r. La seule différence entre notre coefficient de symétrie s et le coefficient de corrélation r (naturellement, outre la signification des x et des y) consiste dans *la nature de la moyenne* qui forme le dénominateur du coefficient : tandis que pour r on prend la moyenne *géométrique* de $\Sigma\, x^2$ et de $\Sigma\, y^2$ $\left(\text{soit : } \sqrt{\Sigma\, x^2\, \Sigma\, y^2}\right)$, nous prenons ici la moyenne *arithmétique* de ces expressions $\left(\dfrac{\Sigma\, x^2 + \Sigma\, y^2}{2}\right)$. (*)

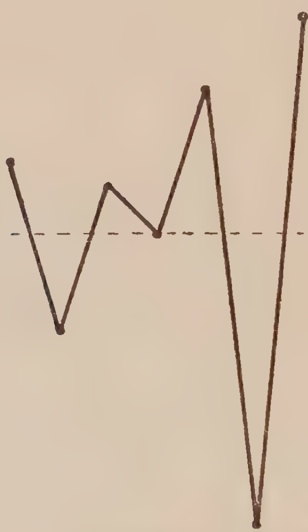
Cette différence nous paraît pleinement justifiée par la différence de nature des deux phénomènes. La corrélation (en ligne droite), pour être parfaite, ne demande pas nécessairement que les x soient

(*) La formule de notre coefficient de symétrie s coïncide avec l'*indice quadratique de corrélation entre les écarts* (absolus) signalé par Gini dans ses *Indici di concordanza* (Atti del « R. Istituto Veneto di Scienze, Lettere ed Arti », 1915-16, pp. 1440-1441).

égaux aux y, il suffit qu'ils soient *proportionnels* ; le coefficient r doit donc marquer la perfection ($= 1$) si $y = xt$, t étant une grandeur constante quelconque. Le coefficient r satisfait à cette condition. Par contre, pour que la symétrie soit parfaite, il ne suffit pas que les x soient proportionnels aux y, il faut qu'ils leur soient *égaux*. Des séries accusant, par exemple, des écarts comme les suivants :

$$x_1 = +2\,;\ x_2 = -\ 4\,;\ x_3 = +3$$
$$y_1 = +6\,;\ y_2 = -12\,;\ y_3 = +9$$

présenteraient une corrélation parfaite ($y = 3x$ ou $t = 3$) et auraient pour coefficient $r = +1$, mais les mêmes écarts marqueraient une série fort éloignée de la parfaite symétrie, comme on peut le voir d'après le graphique ci-contre.



Graphique IX.

Aussi notre coefficient de symétrie s ne devient-il pas égal à 1 dans le cas de variations proportionnelles mais inégales des x et des y. En effet, mettant $x = xt$, le coefficient de symétrie devient :

$$s = \frac{\Sigma xy}{\dfrac{\Sigma x^2 + \Sigma y^2}{2}} = \frac{t\,\Sigma x^2}{\dfrac{\Sigma x^2\,(1 + t^2)}{2}} = \frac{2\,t}{1 + t^2}\,. \qquad [99]$$

Cette expression ne devient $= 1$ que si $2t = 1 + t^2$, soit si $1 + t^2 - 2t = 0$, c'est-à-dire si $(1 - t)^2 = 0$, soit si $t = 1$ ou, ce qui revient au même, si $y = x$. C. Q. F. D.

Dans l'exemple particulier de variations proportionnelles que nous venons de voir (représenté aussi sur le graphique IX), où r donne 1, notre coefficient de symétrie est :

$$s = \frac{2 \cdot 6 + 5 \cdot 15 + 3 \cdot 9}{\frac{(4 + 25 + 9) + (36 + 225 + 81)}{2}} = \frac{114}{190} = 0{,}60\,.$$

(On aurait pu, d'ailleurs, prévoir ce résultat d'après la formule [99] en substituant à t sa valeur 3 ; en effet : $\frac{2 \cdot 3}{1 + 3^2} = \frac{6}{10} = 0{,}60$).

§ 66. — La formule du coefficient de symétrie s telle que nous l'avons écrite plus haut [98] avait pour but surtout de faire mieux ressortir ce que ce coefficient a d'analogue avec le coefficient de corrélation r et en quoi ils diffèrent entre eux. Mais il va de soi que la formule [98] peut être simplifiée.

En effet, l'expression $\Sigma x^2 + \Sigma y^2$ signifie la somme des carrés de tous les écarts, tant gauches que droits ; si nous reprenons le système de notation employé plus haut (§ 39), désignant les écarts en général par E, les écarts situés à gauche du centre de symétrie par E' et les écarts droits par E'', nous pouvons donc écrire :

$$s = \frac{2\,\Sigma\, x\, y}{\Sigma x^2 + \Sigma y^2} = \frac{2\,\Sigma\, E'\, E''}{\Sigma\, E^2}\,. \qquad [100]$$

c'est-à-dire que le *coefficient de symétrie* s *est égal au double de la somme algébrique des produits des écarts des termes équidistants divisée par la somme des carrés de tous les écarts.*

Par E nous désignons toutes sortes d'écarts : moyennaux, centraux et successifs. Pour être explicites, marquons encore la formule du coefficient pour chacune de ces trois espèces de symétrie :

Pour la symétrie des écarts moyennaux (d) :

$$s_d = \frac{2\,\Sigma\, d'\, d''}{\Sigma\, d^2} \qquad [101]$$

Pour la symétrie des écarts successifs (e) :

$$s_e = \frac{2 \sum e' e''}{\sum e^2} \qquad [102]$$

Pour la symétrie des écarts centraux (η) :

$$s_\eta = \frac{2 \sum \eta' \eta''}{\sum \eta^2} . \qquad [103]$$

§ 67. — Avant de montrer par des exemples le calcul de nos coefficients de symétrie, il ne sera peut-être pas inutile de bien préciser un point spécial de ce calcul sur lequel nous avons déjà rapidement attiré l'attention (§ 35 et § 40).

Dans toute série, périodique ou non-périodique, le nombre des écarts moyennaux est égal à celui des termes (n). Il n'en est pas tout-à-fait de même pour les écarts successifs et centraux. Dans les séries périodiques (typiques), le dernier terme, il est vrai, est encore dans la réalité suivi d'un écart successif, par conséquent, dans ces séries il y a aussi autant d'écarts successifs et centraux que de termes (n). Mais dans les séries *non*-périodiques, le dernier terme n'est suivi d'aucun écart successif ; dans ces séries, le nombre des écarts successifs et centraux (les écarts centraux ne sont au fond que les sommes successives des écarts successifs) est ainsi égal au nombre des termes moins 1 (soit $= n - 1$).

Il s'ensuit que dans les séries *non*-périodiques, si le nombre des termes est pair ($2\,m$), le nombre des écarts moyennaux l'est aussi, mais celui des écarts successifs et centraux est impair ($2\,m - 1$), et vice versa : quand le nombre des termes (et des écarts moyennaux) est impair ($2\,m + 1$), celui des écarts successifs et centraux est pair ($2\,m$).

Il s'ensuit encore qu'en cas de symétrie parfaite, le centre de symétrie des écarts *moyennaux* doit coïncider avec le *terme central* de la série si le nombre des termes est *impair* et qu'il doit se placer *entre* les deux termes « centraux » si le nombre des termes est *pair :* alors seulement on aura autant d'écarts à droite qu'à gauche du centre de symétrie. Il en est de même des écarts *successifs et centraux* dans des séries *non-périodiques* : si le nombre des termes est impair ($n = 2\,m + 1$) le nombre de ces écarts est pair ($2\,m$), dont autant (m) à gauche qu'à droite du terme central : c'est donc ce terme, le ($m + 1$)-ième, qui

forme alors le centre de symétrie parfaite des écarts successifs (et centraux) ; si le nombre des termes est pair ($n = 2\,m$) et celui de ces écarts est impair ($2\,m - 1$), il faut donc (pour qu'il y ait autant d'écarts à droite qu'à gauche) que le centre de symétrie parfaite se place entre les deux termes centraux, entre le m-ième et le ($m + 1$)-ième. — Mais pour les séries *périodiques*, qui, elles, ont autant d'écarts successifs et centraux que de termes, c'est juste le *contraire* qui est vrai : si le nombre des termes est *impair* ($2\,m + 1$), celui des écarts successifs (et centraux) l'est également et, pour qu'il y ait autant d'écarts à droite qu'à gauche, le centre de symétrie doit se placer au milieu de l'écart médian, c'est-à-dire *entre* les deux termes centraux ; par contre, si le nombre des termes est *pair* ($2m$) ainsi que celui des écarts successifs et centraux, il faut (pour qu'il y ait symétrie parfaite, pour qu'il y ait m écarts à gauche et autant à droite) que le centre de symétrie coïncide avec un *terme* de la série, avec son ($m + 1$)-ième terme (en commençant d'ailleurs la série par un terme voulu).

Dans le cas de symétrie partielle, il faut naturellement aussi tenir compte de ces différences entre séries ordinaires et périodiques pour fixer correctement les écarts successifs et centraux et pour calculer leur degré de symétrie (ou de dissymétrie). *Pour les séries périodiques, il ne faut surtout pas oublier l'écart successif qui suit le dernier terme de la série* (ou, ce qui revient au même, qui précède le premier terme), c'est-à-dire qui sépare le dernier terme d'une période du premier de la période suivante ; il en est de même de l'écart central qui sépare le centre d'une période du premier terme de la période suivante (ou du dernier terme de la période précédente).

Quelques exemples :

1. Le nombre des termes est impair :

a) Série ordinaire :

v . . . 3, 5, 10, 8, 4		($v_m = 6$)
d . . . — 3, — 1, + 4, + 2, — 2		(le centre de symétrie est le 3me terme : 10)
e . . . — 2, — 5,	— 2, — 4	(*idem*)
η . . . — 7, — 5,	— 2, — 6	(*idem*)

b) Série périodique :

$v \ldots (3, 5, 10, 8, 4)$	($v_m = 6$)
$d \ldots (-3, -1, +4, +2, -2)$	(le centre de symétrie est le 3^me^ terme : 10)
$e \ldots (-2, -5, +1, -1, -4, -1)$	(le centre de symétrie se place entre le 3^me^ et le 4^me^ termes égaux resp. à 10 et à 8, et doit être mis = 9)
$\eta \ldots (-6, -4, +1, -1, -5, -6)$	(*idem*)

2. Le nombre des termes est pair :

a) Série ordinaire :

$v \ldots 3, 7, 9, 5$	($v_m = 6$)
$d \ldots -3, +1, +3, -1$	(le centre de symétrie se place entre le 2^me^ et le 3^me^ termes)
$e \ldots -4, -1, +1, -4$	(*idem* ; le centre de symétrie, placé entre 7 et 9, doit être mis = 8)
$\eta \ldots -5, -1, +1, -3$	(*idem*)

b) Série périodique :

$v \ldots (3, 7, 9, 5)$	($v_m = 6$)
$d \ldots (-3, +1, +3, -1)$	(le centre de symétrie se place entre le 2^me^ et les 3^me^ termes)
$e \ldots (-4, -2, -4, -2)$	(le centre de symétrie est le 3^me^ terme : 9)
$\eta \ldots (-6, -2, -4, -6)$	(*idem*).

Les observations ci-dessus ont pour objet de faire placer correctement le point médian de la série des écarts qui est généralement aussi leur centre de symétrie. Mais, comme nous l'avons déjà vu (§ 22, 1° et § 41, dernière partie), il existe des cas particuliers où la série se montre par rapport à son point médian moins symétrique que par rapport à quelque autre terme de la série ; dans des cas pareils,

le degré de symétrie doit être fixé en prenant comme centre de symétrie le point de la série marquant la dissymétrie minima (ou la symétrie maxima) ; là encore il ne faut jamais perdre de vue le fait que les séries périodiques ont un nombre d'écarts successifs et centraux supérieur de 1 à celui des séries non-périodiques ayant le même nombre de termes.

§ 68. — Prenons maintenant quelques exemples de calcul pratique des coefficients de symétrie, et afin de pouvoir mieux comparer la signification concrète du coefficient de symétrie s avec celle de l'indice de dissymétrie D, calculons s pour quelques unes des séries pour lesquelles nous avons déjà établi l'indice D.

Prenons par exemple les quatre séries du § 63 ; nous obtenons pour elles les coefficients de symétrie suivants (pour faciliter la comparaison, nous donnons dans des crochets [ ] les indices de dissymétrie trouvés plus haut) :

Série I :

$$s_1 = \frac{2\,(2 \,.\, 3 + 1 \,.\, 3 + 1 \,.\, 2 + 3 \,.\, 5)}{2^2 + 3^2 + 1^2 + 5^2 + 3^2 + 1^2 + 2^3 + 3^2} = \frac{2.26}{62} =$$

$$= 0{,}84 \,;\ [D_1 = 0{,}30\ (-1)]$$

Série II :

$$s_2 = \frac{2\,(2 \,.\, 3 + 2 \,.\, 3 - 1 \,.\, 1 + 3 \,.\, 5)}{62} = \frac{2 \,.\, 26}{62} = 0{,}84 \,;\ [D_2 = 0{,}30\ (-0{,}90)]$$

Série III :

$$s_3 = \frac{2\,(-2 \,.\, 2 + 3 \,.\, 3 + 2 \,.\, 1 + 3 \,.\, 4)}{56} = \frac{2 \,.\, 19}{56} = 0{,}68 \,;\ [D_3 = 0{,}30\ (-0{,}80)]$$

Série IV :

$$s_4 = \frac{2\,(2 \,.\, 2 - 3 \,.\, 3 + 2 \,.\, 2 + 3 \,.\, 3)}{52} = \frac{2 \,.\, 8}{52} = 0{,}31 \,;\ [D_4 = 0{,}30\ (-0{,}70)]$$

On remarque que d'une façon générale, il y a une grande concordance entre les conclusions basées sur le coefficient de symétrie s et celles basées sur les *deux* indices de dissymétrie. Mais on voit aussi combien il serait nécessaire de prendre en considération à la fois les

deux indices de dissymétrie, combien fallacieuses peuvent être les indications d'un seul d'entre eux. On voit également combien le coefficient de symétrie se montre plus synthétique et plus expressif. On remarquera en outre (en comparant les résultats pour les séries II, III et IV) que le coefficient de symétrie s, basé sur le *produit* des écarts, se trouve plus fortement influencé par les grandes déviations de la symétrie que les indices de dissymétrie D basés sur l'addition (ou la différence) de ces déviations.

Prenons encore, à titre d'illustration, l'exemple des séries III et IV du § 41.

Série III :

$$s_d = \frac{2\,(-7 \cdot 5 + 5 \cdot 2 + 7 \cdot 2)}{7^2 + 2^2 + 5^2 + 7^2 + 5^2 + 2^2} = \frac{2\,(-11)}{78 \cdot 2} = -0{,}14\,;$$

$$[D_d = -0{,}64\,(+0{,}71)]\,.$$

Série IV (*) :

$$s_d = \frac{2\,(-5 \cdot 7 + 7 \cdot 2 + 2 \cdot 0 + 5)}{78 \cdot 2} = \frac{-21}{78} = -0{,}27\,;$$

$$[D_d = -0{,}64\,(+0{,}86)]\,.$$

Comme le montrent les *indices de dissymétrie D*, les deux séries s'écartent beaucoup et de la symétrie positive et de la symétrie négative, moins toutefois de la négative que de la positive ; il en résulte, pour les deux séries, de très faibles *coefficients de symétrie négative*. Cependant, pour la première de ces séries (III), où la différence des indices de dissymétrie positive et de dissymétrie négative est plus faible, le coefficient de symétrie négative est aussi plus faible ; pour la dernière de ces séries (IV), où la différence entre l'indice de dissymétrie positive et celui de la dissymétrie négative est plus sensible, le coefficient de symétrie (négative) est aussi plus important.

(*) Il s'agit de la modalité spéciale signalée plus haut : on cherche le degré de symétrie des écarts *moyennaux* alors que le terme central marque lui-même un écart de la moyenne. Comme pour les indices de dissymétrie, il est juste de tenir compte de cet écart pour le numérateur comme pour le dénominateur de notre coefficient ; le résultat final devient de ce fait plutôt approximatif, il est cependant d'autant plus strict que l'écart accusé par le terme central est moins important relativement au nombre et à l'importance des autres écarts.

Il est vrai que la connexité entre les indices de dissymétrie D et le coefficient de symétrie n'est pas absolue (surtout par suite du poids différent des grandes déviations dans le calcul de nos indices et de notre coefficient) ; comme nous voyons cependant d'après ces exemples, elle est assez générale. Les détails fournis par les indices de dissymétrie peuvent, en outre, compléter utilement le coefficient de symétrie qui est plus synthétique. Cette circonstance pourrait, peut-être, justifier dans certains cas l'emploi des indices de dissymétrie D à *côté* du coefficient de symétrie s , mais nous ne croyons pas qu'ils puissent le remplacer.

CHAPITRE X.

Application du système d'indices examiné dans ce travail à l'étude de quelques périodicités réelles.

§ 69. — Afin de voir d'une façon concrète dans quelle mesure les divers indices examinés plus haut facilitent la comparaison de différentes variations périodiques et aident à dégager leurs caractères propres, nous ferons un certain nombre d'applications de ces indices à l'étude de certaines variations périodiques réelles, observées dans plusieurs pays et pour des phénomènes de nature différente. Les séries réelles elles-mêmes se trouvent indiquées dans notre annexe II, à la fin de ce travail ; ici, nous donnerons les indices de la périodicité de ces séries et leur analyse sommaire. Pour chaque série périodique, nous relèverons notamment :

1º La moyenne arithmétique de la série (v_m).

2º La partie de la période pendant laquelle le phénomène se trouve au-dessus (et au-dessous) de la moyenne.

3º Le nombre des sommets.

4º La valeur et le moment des variations extrêmes (du maximum et du minimum principaux).

5º L'amplitude des variations (α) , absolue et relative.

6º L'écartement de la moyenne (d_m et σ), absolu et relatif.

7º L'écartement des termes successifs (e_m et ς), absolu e trelatif.

8º Les coefficients de concentration (k_n et c_n) et de nivellement (ν_n).

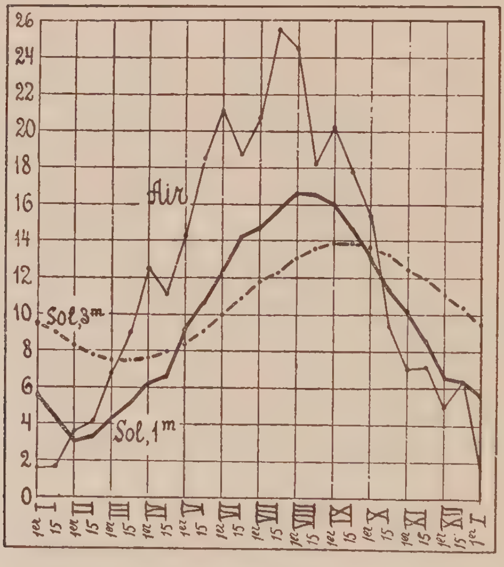
9° Les coefficients de précipitation (p_n) et de gradualité (g_n).
10° Les coefficients de symétrie (s) :

a) des écarts moyennaux (s_d),
b) des écarts successifs (s_e) et
c) des écarts centraux (s_η).

Dans certains cas, nous procèderons encore à la péréquation des séries données par la sinusoïde.

En outre, dans la première application concrète, nous ferons usage, à titre d'illustration, de nos indices inachevés de la concentration (K et C_n) et de la précipitation (P_n) ainsi que des indices de la dissymétrie (D) ; pour la première application, nous donnons aussi, dans l'annexe II, des modèles du calcul de nos indices et coefficients.

§ 70. — Nos premières applications se rapporteront au domaine *géophysique*. Le *Statistisches Jahrbuch der Stadt Berlin* contient beaucoup de renseignements intéressants sur les variations périodiques



GRAPHIQUE X.

de phénomènes de cet ordre à Berlin. Nous lui avons emprunté, un peu au hasard, les données ayant servi à former le tableau I de l'*Annexe* II et montrant, pour chaque date indiquée, la *température moyenne de l'air atmosphérique* ainsi que celle du *sol* à *un mètre et à trois mètres*

de profondeur, d'après les observations faites, le 1er et le 15 de chaque mois, à sept endroits de la ville de Berlin, pendant les années 1910-1914 (les termes des trois séries de ce tableau présentent des moyennes arithmétiques des températures observées au même moment de l'année pendant les cinq années en question).

Nous examinerons, d'un côté, les données observées le 15 de chaque mois formant ainsi des séries périodiques à 12 termes et, de l'autre les données observées le 1er et le 15 de chaque mois, constituant pour les mêmes phénomènes des séries périodiques à 24 termes. De cette façon, nous pourrons comparer les variations périodiques des trois phénomènes en question (température de l'air, du sol à 1^{m} de profondeur et du sol à 3^{m} de profondeur) et nous pourrons voir aussi ce que la plus grande fréquence des observations apporte à une connaissance plus stricte de ces phénomènes. Afin de pouvoir mieux voir dans quelle mesure les divers indices reflètent la réalité, nous donnons aussi une représentation graphique des variations des trois phénomènes.

Si nous examinons maintenant ces six séries périodiques typiques (dont trois à 12 termes et trois à 24) d'après le schéma indiqué au § précédent, nous obtenons les résultats suivants :

Caractères de la périodicité annuelle des variations de la température : de l'air atmosphérique, du sol à 1^m de profondeur et du sol à 3^m de profondeur, à Berlin (1910-1914).

	D'après les données du 15 de chaque mois (12 termes)			D'après les données du 1^er et du 15 de chaque mois (24 termes)		
	Air atmosphérique	Sol profondeur 1^m	Sol profondeur 3^m	Air atmosphérique	Sol profondeur 1^m	Sol profondeur 3^m
1. Température moyenne de l'année (v_m)	12°,3	9°,8	10°,65	12°,5	9°,8	10°,63
2. Partie de l'année ayant une température :						
supérieure à la moyenne	0,42	0,50	0,50	0,46	0,50	0,50
inférieure à la moyenne	0,58	0,50	0,50	0,50	0,50	0,50
3. Nombre des sommets	1	1	1	6	1	1
4. Température maxima :						
valeur (v_{max})	25°,5	16°,5	13°,9	25°,5	16°,6	13°,9
moment de l'année	15/VII	15/VIII	15/IX	15/VII	1/VIII	1-15/IX
Température minima :						
valeur (v_{min})	1°,7	3°,3	7°,5	1°,6	3°,1	7°.5
moment de l'année	15/I	15/II	15/III	1 (15)/I	1/II	1-15/ II
5. Amplitude des variations :						
absolue (α)	23°,8	13°,2	6°,4	23°,9	13°,5	6°,4
relative : $\frac{\alpha}{v_m}$	1,93	1,35	0,60	1,91	1,38	0,60
» $\frac{\alpha}{2\sigma}$	1,72	1,47	1,42	1,66	1,50	1,41
6. Ecartement de la moyenne :						
absolu : écart arithmétique moyen (d_m)	6°,19	4°,06	2°,02	6°,43	4°,03	2°,04
absolu : écart quadratique moyen (σ)	6°,94	4°,50	2°,26	7°,22	4°,50	2°,27
relatif : $\frac{d_m}{v_m}$	0,50	0,41	0,19	0,51	0,41	0,19
» $\frac{\sigma}{v_m}$	0,56	0,46	0,21	0,58	0,46	0,21

	D'après les données du 15 de chaque mois (12 termes)			D'après les données du 1er et du 15 de chaque mois (24 termes)		
	Air atmosphérique	Sol profondeur 1m	Sol profondeur 3m	Air atmosphérique	Sol profondeur 1m	Sol profondeur 3m
7. Écartement des termes successifs :						
absolu : écart successif arithmétique moyen (e_m)	4°,00	2°,20	1°,07	2°,60	1°,13	0°,53
absolu : écart successif quadratique moyen (ς)	4°,88	2°,40	1°,17	3°,06	1°,28	0°,60
relatif : $\frac{ς}{v_m}$	0,40	0.24	0,11	0,24	0,13	0,06
» $\frac{ς}{\sigma}$	0,71	0,53	0.52	0,42	0,28	0,26
8. Concentration et nivellement :						
coefficient de concentration k_n .	0,22	0,25	0,27	0,22	0,23	0,23
coefficient de concentration c_n .	0,08	0,08	0,08	0.05	0,05	0,05
coefficient de nivellement v_n . .	0,82	0,83	0,82	0,85	0,86	0,86
9. Coefficients de :						
précipitation p_n	0,15	0,06	0,06	— (*)	0,06	0,05
gradualité g_n	0,70	0,87	0,85	— (*)	0.83	0,85
10. Coefficients de symétrie s :						
des écarts moyennaux (s_d) . . .	−0,951	−0,958	−0,994	−0,963	−0,993	0,992
(centre de symétrie)	15/x	1/xi	15/xii	15/x	8/viii	8/ix
des écarts successifs (s_e)	0,923	0,872	0,979	0,680	0,873	0,973
(centre de symétrie)	15/vii	15/viii	15/ix	22/vii	8/viii	8/ix
des écarts centraux (s_η)	0,996	0,986	0,999	0,989	0,998	9,997
(centre de symétrie)	15/vii	15/viii	15/ix	22/vii	8/viii	8/ix

(*) La série n'étant pas complètement monocéphale, nous avons renoncé à calculer les coefficients de précipitation (p_n) et de graduatlité (g_n).

§ 71. — Relevons les principaux points de ce tableau :

1° La température annuelle *moyenne* de l'air atmosphérique est plus élevée que celle du sol, observé à un et à trois mètres de pro-

fondeur ; la différence de température moyenne entre l'air et le sol à 1^m de profondeur est de 2°,7 selon les observations bi-mensuelles et de 2°,5 selon les observations mensuelles. Cependant, plus en profondeur, la température remonte ; à 3 mètres de profondeur, elle est sensiblement plus élevée qu'à 1 mètre, à savoir de 0°,83 (respectivement de 0°,85) pour une différence de profondeur de 2 mètres. (Nous ne nous arrêterons pas sur l'explication et les suggestions que comporte cette constatation ; cela nous amènerait loin de l'objet propre de notre étude).

2° La température du sol est au-dessus de la moyenne durant exactement une moitié de l'année, elle est au-dessous de la moyenne durant l'autre moitié. Celle de l'air semble être au-dessous de la moyenne un peu plus longtemps qu'elle n'est au-dessus d'elle.

3° La périodicité annuelle de la température du *sol*, observée une fois par mois ou par quinzaine, à un mètre comme à trois de profondeur, présente une courbe *monocéphale*. La périodicité annuelle de la température de *l'air* est au fond également monocéphalique, comme on le voit d'après la série à 12 termes ; cependant, la température de l'air est soumise à un nombre important de fluctuations secondaires de courte durée qui, dans la réalité, se greffent sur la périodicité monocéphale annuelle. Par conséquent, si l'on observe cette température à des petits intervalles, à l'intérieur desquels l'influence de la périodicité annuelle se fait peu sentir, on y constate des oscillations notables provoquées par ces circonstances secondaires ; avec 24 observations par an, on trouve, comme nous le voyons, 6 maxima (leur nombre augmenterait sans doute encore beaucoup si l'on réduisait davantage les intervalles entre les mensurations). La monocéphalie persistante des variations de la température du sol est d'autant plus significative (il est toutefois probable qu'en réduisant encore plus les intervalles entre les observations, on finirait ici aussi par découvrir un certain nombre de petits sommets dus à l'action de causes secondaires).

4° Par rapport à la révolution de la terre autour du soleil, la périodicité annuelle de la température est pour l'air atmosphérique en retard d'environ 3 semaines, le maximum de cette température tombant au 15 juillet (soit environ 3 semaines après le solstice d'été) et le minimum au 15 janvier (environ 3 semaines après le solstice d'hiver) ; pour le sol à 1 mètre de profondeur, le retard est d'environ 6 semaines, et pour le sol à 3 mètres de profondeur, d'environ 11 semaines. C'est ainsi que le maximum de température est atteint par le sol à cette dernière profondeur environ deux semaines avant l'équinoxe

d'automne (vers le 7 septembre) et le minimum environ deux semaines avant l'équinoxe de printemps (vers le 7 mars).

5° *L'intensité des fluctuations saisonnières*, d'après les observations mensuelles comme d'après les observations bi-mensuelles, que l'on envisage *l'amplitude* des variations, *l'écartement de la moyenne* ou l'écartement des termes *successifs* est pour l'air beaucoup plus forte que pour le sol, et pour le sol à une profondeur moindre (1^m) beaucoup plus forte qu'à une profondeur plus grande (3^m). L'écart moyennal moyen, arithmétique (d_m) ou quadratique (σ), que l'on envisage les mensurations mensuelles ou bi-mensuelles, se montre pour l'air atmosphérique *trois fois plus fort* que pour le sol à 3^m de profondeur (d'après les observations bi-mensuelles, par exemple, d_m est égal respectivement à 6°,4 , à 4°,3 et à 2°,0 ; σ est respectivement égal à 7°,2 à 4°,5 et à 2°,3). L'amplitude des variations (α) est pour le sol à 3 mètres de profondeur environ deux fois plus faible qu'à un mètre de profondeur et environ quatre fois plus faible que pour l'air atmosphérique (d'après les observations bi-mensuelles, α est égal respectivement à 23°,9 , à 13°,5 et à 6°,4). L'écart successif moyen, arithmétique (e_m) ou quadratique (ς), est pour le sol à trois mètres 4 à 5 fois plus faible que pour l'air atmosphérique (pour les séries à 24 termes, e_m est égal respectivement à 2°,6 , à 1°,1 et à 0°,5 ; ς est respectivement de 3°,1 , 1°,3 et 0°,6). Même *relativement*, les fluctuations saisonnières sont pour l'air beaucoup plus fortes que pour le sol, et pour le sol même ces fluctuations sont à une profondeur moindre encore considérablement plus fortes qu'à une profondeur plus grande. Ainsi, par rapport à la température annuelle moyenne (v_m), l'amplitude des variations, l'écart moyennal quadratique moyen et l'écart successif quadratique moyen s'expriment, d'après les observations bi-mensuelles, par les chiffres suivants :

	$\frac{\alpha}{v_m}$	$\frac{\sigma}{v_m}$	$\frac{\varsigma}{v_m}$
Air atmosphérique	1,91	0,58	0,24
Sol à 1^m de profondeur . . .	1,38	0,46	0,13
Sol, à 3^m de profondeur . .	0,60	0,21	0,06

L'amplitude des variations constitue ainsi près du double (1,91) de la température annuelle moyenne de l'air atmosphérique, elle dépasse de plus d'un tiers (1,38) la moyenne annuelle de la température du sol à un mètre de profondeur et elle n'est que 60 p. 100 de la température annuelle moyenne pour le sol à trois mètres de profondeur ;

même relativement, l'amplitude des variations est ainsi pour le sol à 3^m de profondeur trois fois plus faible que pour l'air atmosphérique. La proportion est, comme on le voit, à peu près la même pour l'écartement de la moyenne $\left(\frac{\sigma}{v_m}\right)$. L'écartement successif $\left(\frac{\varsigma}{v_m}\right)$ est, même relativement, pour le sol à 3^m de profondeur environ deux fois plus faible qu'à un mètre de profondeur et quatre fois plus réduit que pour l'air atmosphérique (respectivement, 0,06, 0,13 et 0,24). Et il en est de même d'après les observations mensuelles.

6° Quant à la *répartition* de l'écartement moyennal et de l'écartement successif sur les différents termes des séries considérées (c'est-à-dire quant au degré de la *concentration* et au degré de la *précipitation* de leurs variations), on peut faire les constatations suivantes : toutes ces variations ont une très faible concentration (d'après les observations bi-mensuelles, le coefficient de concentration k_{24} est pour les trois phénomènes égal à 0,22 ou 0,23, le coefficient de concentration c_{24} n'est que 0,05) et une très faible précipitation (p_{24} étant pour le sol égal à 0,05 ou 0,06) ; ce sont des variations très égales, fort nivelées et graduelles, formant des courbes à allure calme, ondulatoire ; ce n'est que la température de l'air atmosphérique qui accuse une précipitation de quelque importance ($p_{12} = 0,15$ pour les séries à 12 termes), la concentration demeurant pour l'air tout aussi faible que pour la température du sol (On aura, en outre remarqué combien dans ces séries réelles à très faible concentration, le coefficient k_n se montre cependant plus élevé que c_n ; en partie, au moins, c'était d'ailleurs à prévoir étant donné le caractère des deux coefficients examiné plus haut, § 62).

7° Enfin, la *symétrie* des variations de chacun des trois phénomènes considérés est tout-à-fait remarquable. Elle est remarquable, d'abord, par son degré extrêmement élevé qui touche pour ainsi dire à la perfection (nous avons calculé les coefficients de symétrie avec trois décimales, car en nous bornant à deux, nous aurions obtenu dans certains cas, notamment pour la symétrie des écarts centraux, le coefficient 1,00, symbole de la perfection). La symétrie de ces variations est remarquable encore parce qu'elle est extrêment élevée sous tous les rapports : comme symétrie des écarts moyennaux, des écarts successifs et des écarts centraux. (Nous avons d'ailleurs vu plus haut, au § 20, qu'en général, dans le cas de symétrie parfaite de l'une des trois espèces d'écarts, les deux autres accusent une symétrie éga-

lement parfaite). La symétrie est extraordinairement élevée pour les trois phénomènes : la température de l'air, la température du sol à 1 mètre et celle du sol à 3 mètres de profondeur, mais ce sont les variations de cette dernière température qui, sous tous les rapports, sont d'une symétrie tout-à-fait étonnante, allant de 0,97 à 1,00.

Notons encore que ces variations forment des séries très symétriques non seulement par rapport aux centres de symétrie indiqués sur notre tableau, mais dans une mesure égale ou presque égale aussi par rapport aux points opposés à ceux que nous avons pris comme centres ; ce sont, en d'autres termes, des séries très *bi-symétriques*. Ces variations, et surtout celles de la température du sol à 3 mètres de profondeur, sont même très symétriques encore par rapport aux points se trouvant à égale distance (ou presque) de ces deux centres de symétrie, à peu près également légitimes ; ce sont donc des séries très *tétra-symétriques*. Ainsi, à prendre les observations bi-mensuelles (séries à 24 termes), on peut constater que les écarts de la moyenne annuelle sont pour l'air atmosphérique très symétriques non seulement par rapport à la température du 15 octobre et du 15 avril, mais aussi par rapport à celles du 8 juillet et du 8 janvier ($s_d = 0{,}904$) ; les mêmes écarts sont pour le sol à 1 mètre de profondeur excessivement symétriques non seulement par rapport au 8 août et au 8 février, mais presque dans la même mesure aussi par rapport au 8 mai et au 8 novembre pris comme centres de symétrie ($s_d = -0{,}988$) ; pour le sol observé à 3 mètres de profondeur, la symétrie est quasi-parfaite non seulement par rapport au 8 septembre et au 8 mars, mais aussi par rapport au 8 juin et au 8 décembre ($s_d = -0{,}992$). Et ainsi de suite, pour les autres formes de symétrie.

§ 72. — Afin d'être plus complets et de nous rendre compte plus concrètement de la valeur des divers indices dégagés dans la présente étude, nous donnerons encore les indices inachevés de la concentration et de la précipitation ainsi que les indices de la dissymétrie, pour les six séries de variations que nous venons d'examiner :

	D'après les données du 15 de chaque mois (12 termes)			D'après les données du 1er et du 15 de chaque mois (24 termes)		
	Air atmosphérique	Sol profondeur 1^m	Sol profondeur 3^m	Air atmosphérique	Sol profondeur 1^m	Sol profondeur 3^m
Indices de concentration :						
$\frac{\sigma}{d_m}$	1,12	1,11	1,12	1,12	1,12	1,11
C_n	1,25	1,23	1,25	1,26	1,25	1,23
K	0,36	0,41	0.45	0,41	0,42	0,43
Indices de nivellement :						
$\frac{d_m}{\sigma}$	0,89	0,90	0,89	0,89	0,90	0,90
$N_n \left(= \frac{d_m}{\sigma}\right)^2$	0,79	0,81	0,79	0,79	0,80	0,81
Indices de précipitation :						
$\frac{\varsigma}{\alpha}$	0,21	0,18	0,18	— (*)	0,09	0,09
$\frac{\varsigma}{e_m}$	1,22	1,09	1,09	— (*)	1,14	0,13
P_n	1,49	1,19	1,19	— (*)	1,30	1,27
Indices de gradualité :						
$\frac{\alpha}{\varsigma}$	4,88	5,50	5,47	— (*)	10,55	10,67
$\frac{e_m}{\varsigma}$	0,82	0,92	0,91	— (*)	0,88	0,89
$G_n \left(= \frac{e_m}{\varsigma}\right)^2$	0,67	0,85	0,83	— (*)	0,77	0,79
Indices de dissymétrie (D) :						
des écarts moyennaux (D_d)	— 0,16 (+ 1,00)	— 0,13 (+ 1,00)	— 0,06 (+ 1,00)	— 0,14 (+ 1,00)	+ 0,05 (— 1,00)	+ 0,06 (— 1,00)
(centre de symétrie) .	15/X	1/XI	15/XII	15/X	8/VIII	8/IX
des écarts successifs (D_e)	+ 0,17 (— 1,00)	+ 0,25 (— 1,00)	+ 0,09 (— 1,00)	+ 0,40 (— 0,79)	+ 0,13 (— 0,99)	+ 0,09 (— 1,00)
(centre de symétrie) .	15/VII	15/VIII	15/IX	22/VII	8/VIII	8/IX
des écarts centraux (D_η)	+ 0,04 (— 1,00)	+ 0,07 (— 1,00)	+ 0,09 (— 1,00)	+ 0,07 (— 1,00)	+ 0,03 (— 1,00)	+ 0,03 (— 1,00)
(centre de symétrie) .	15/VII	15/VIII	15/IX	15/VII	8/VIII	8/IX

(*) La série n'étant pas complètement monocéphale, nous n'avons pas calculé cet indice.

Qu'est-ce que nous apprennent les indices de concentration et de nivellement, de précipitation et de gradualité ? Ils montrent que les variations, mensuelles ou bi-mensuelles de la température de l'air atmosphérique, du sol à 1^m de profondeur et à 3^m de profondeur ont à peu près la même concentration ou le même nivellement ; ils montrent qu'elles sont pour le sol moins précipitées, plus graduelles que pour l'air atmosphérique. Mais ils ne nous disent pas si ces caractères (précipitation, concentration, etc.) sont pour les phénomènes considérés faibles ou forts ; ce n'est que par l'expérience d'autres séries périodiques qu'on pourrait se faire une idée, d'ailleurs assez vague, de l'intensité des caractères qui nous intéressent ici. Nous pensons donc que ces indices inachevés n'ajoutent rien aux coefficients correspondants et que l'on peut, par conséquent, se dispenser de leur calcul une fois les coefficients établis.

Plus nets sont les indices de dissymétrie (qui marquent toute déviation de la symétrie en fraction de la déviation maxima) ; les indications qu'ils fournissent concordent, en outre, avec celles données par nos coefficients de symétrie. Ils ajoutent quelquefois aussi à la compréhension des coefficients ; ainsi nous voyons que la symétrie relativement peu élevée des écarts successifs (bi-mensuels) de la température de l'air que nous avons trouvée au § précédent ($s_e = 0{,}68$), provient en grande partie du fait que les écarts équidistants du centre de symétrie se font parfois en sens opposés [$D_e = +0\ ,\ 40\ (-0{,}79)$]. D'une façon générale pourtant, l'utilité de ces indices de dissymétrie se montre aussi fort restreinte quand on a déjà établi les coefficients de symétrie. Dans la suite nous nous bornerons donc à calculer seulement les coefficients, laissant de côté les indices correspondants.

§ 73. — Quoi qu'il en soit, les périodicités annuelles des variations de la température de l'air atmosphérique et du sol, notamment du sol à trois mètres de profondeur, forment comme nous l'avons vu, des séries monocéphales, à faible concentration, d'une allure graduelle, ondulatoire (très peu précipitée) et remarquablement symétriques, et non seulement bi-symétriques, mais tétra-symétriques. Ces séries possèdent ainsi à un haut degré tous les caractères essentiels de la *sinusoïde*. On doit donc pouvoir les représenter par des sinusoïdes avec une approximation bien admissible et sans faire trop de violence à la réalité (Cf. § 26). La courbe de péréquation doit se rapprocher de la courbe empirique tout particulièrement pour la température du sol à 3^m de profondeur dont les variations sont les plus symétriques (notam-

ment pour les écarts successifs) et les plus graduelles. Nous pouvons remarquer en particulier que l'amplitude des variations relativement à l'écartement moyennal, à savoir le rapport, $\frac{\alpha}{2\sigma}$ est pour la température du sol à 3^m de profondeur exactement égal à celui que nous avons établi plus haut [13] pour la sinusoïde. (Ce rapport est, d'après notre tableau du § 70, égal à 1,72 et 1,66 pour l'air atmosphérique, à 1,47 et 1,50 pour le sol à 1^m de profondeur et à 1,42 et 1,41 pour le sol à 3^m de profondeur, selon que l'on envisage les observations mensuelles ou bi-mensuelles ; il est, d'après la formule [13], égal à 1,414 pour la sinusoïde théorique).

Dans ces conditions, la représentation par une sinusoïde (ou par sa formule analytique) n'est pas seulement permise, mais paraît s'imposer, tout particulièrement pour les variations de la température du sol à 3^m de profondeur. Pour pouvoir comparer les variations des trois phénomènes à l'aide de cette méthode, nous l'appliquerons aux variations des trois ; pour la péréquation, nous nous baserons sur les variations bi-mensuelles qui serrent la réalité de plus près ; nous le ferons même pour la température de l'air dont les variations bi-mensuelles présentent un nombre important de sommets, d'ailleurs très secondaires. De cette façon non seulement nous pourrons comparer les trois phénomènes considérés, mais nous verrons aussi mieux les avantages et les inconvénients de la méthode.

§ 74. — Pour opérer la péréquation de nos séries périodiques empiriques par la sinusoïde $y = a + b \sin x$, a doit être pris égal à la moyenne arithmétique de chaque série ($= v_m$) ; les séries comptant chacune 24 termes, les x varieront d'un terme à l'autre de $\frac{360^o}{24}$, c'est-à-dire de 15° en 15°. En outre, la sinusoïde a le plus de chances de se rapprocher le plus de la série empirique si l'on met son point d'origine (0°) au centre de la symétrie négative des écarts moyennaux (autrement dit, *à égale distance entre* les deux centres de la symétric moyennale positive, marquant chacun resp. le *maximum* et le *minimum* de la série. Le point 0° correspondra ainsi au 8 juin ou au 8 décembre pour le sol à 3^m de profondeur, au 8 mai ou au 8 novembre pour le sol à 1^m de profondeur et au 8 (ou 15) avril ou au 8 (ou 15) octobre pour l'air

atmosphérique (*). La température du sol à 3^m de profondeur allant en augmentant pendant les six mois dont le 8 juin est l'exact milieu, tout comme la sinusoïde va en augmentant pour la demi-circonférence dont le milieu est 0^o, le paramètre b sera positif si nous mettons le point 0^o au 8 juin ; si, au contraire, nous mettons le point 0^o au 8 décembre, qui constitue l'exact milieu des six mois pendant lesquels la température du sol à 3^m de profondeur va en diminuant, b sera négatif ; mais dans un cas comme dans l'autre, la valeur absolue de b sera la même. Il en sera de même des dates du 8 mai ou 8 novembre pour le sol à 1^m de profondeur, d'avril ou octobre pour l'air atmosphérique. Pour plus de simplicité, nous prendrons comme point d'origine de la sinusoïde les dates correspondant à b positif. Quant à b lui-même nous le calculons d'après le procédé simplifié indiqué au § 25, formule [10], c'est-à-dire en mettant $b = \dfrac{\Sigma\,(v \sin x)}{12}$.

Pour les trois séries périodiques à 24 termes, nous obtenons ainsi les formules de péréquation suivantes :

Pour la température du sol observée à 3^m de profondeur :

$$y = 10^o{,}6 + 3^o{,}2 \sin x\,, \qquad [104]$$

$10^o{,}6$ étant la température annuelle moyenne du sol à 3^m de profondeur (plus exactement : $10^o{,}63$) et correspondant à la position de la sinusoïde au 8 juin (0^o) et au 8 décembre (180^o).

Pour la température du sol à 1^m de profondeur :

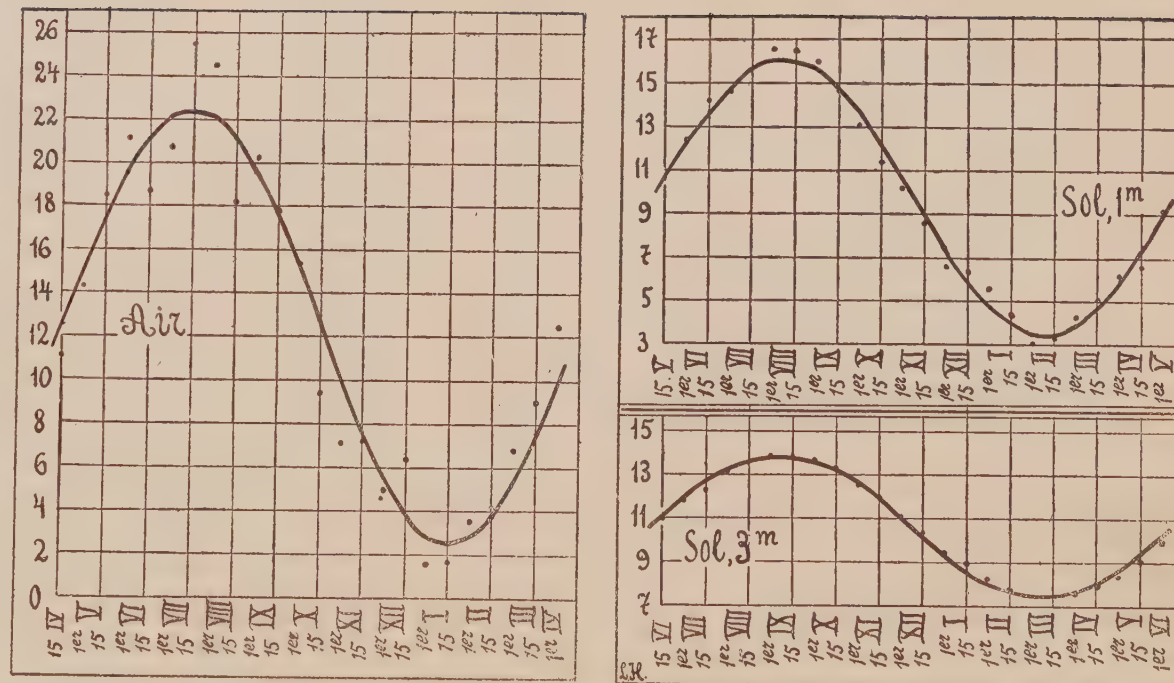
$$y = 9^o{,}8 + 6^o{,}33 \sin x, \qquad [105]$$

$9^o{,}8$ étant la température annuelle moyenne du sol à 1^m de profondeur et correspondant à la hauteur de la sinusoïde au 8 mai (0^o) et au 8 novembre (180^o).

Pour l'air atmosphérique :

$$y = 12^o{,}5 + 9^o{,}733 \sin x\,, \qquad [106]$$

(*) La divergence entre la sinusoïde théorique et la courbe empirique est pour l'air atmosphérique presque exactement la même qu'on mette le point 0^o au 8 ou au 15 des mois indiqués (Voy. les chiffres correspondants dans les deux notes en bas des pages qui suivent).



GRAPHIQUE XI.

(Les points marquent les nombres empiriques pour autant qu'ils s'écartent de la sinusoïde).

Température du sol, à 3^m et à 1^m de profondeur, et température de l'air atmosphérique, à Berlin, d'après les données de 1910-1914 (en degrés Celsius).

Dates	Sol, profondeur 3^m			Sol, profondeur 1^m			Air atmosphérique		
	chiffres réels (v)	chiffres théoriques (y)	$v-y$	chiffres réels (v)	chiffres théoriques (y)	$v-y$	chiffres réels (v)	chiffres théoriques (y)	$v-y$
1er janvier . .	9,5	9,4	+ 0,1	5,6	4,8	+ 0,8	1,6	2,9	— 1,3
15 » . .	9,0	8,7	+ 0,3	4,4	4,0	+ 0,4	1,7	2,6	— 0,9
1er février . .	8,3	8,1	+ 0,2	3,1	3,5	— 0,4	3,6	2,9	+ 0,7
15 » . .	7,8	7,7	+ 0,1	3,3	3,5	— 0,2	4,2	3,9	+ 0,3
1er mars »	7,5	7,5	0	4,3	4,0	+ 0,3	6,8	5,5	+ 1,3
15 » . . .	7,5	7,5	0	5,1	4,8	+ 0,3	9,0	7,6	+ 1,4
1er avril . . .	7,6	7,7	— 0,1	6,2	6,0	+ 0,2	12,5	9,9	+ 2,6
15 » . . .	8,0	8,1	— 0,1	6,6	7,4	— 0,8	11,1	12,5	— 1,4
1er mai	8,4	8,7	— 0,3	9,2	9,0	+ 0,2	14,3	15,1	— 0,8
15 »	9,1	9,4	— 0,3	10,6	10,6	0	18,5	17,4	+ 1,1
1er juin	10,0	10,2	— 0,2	12,4	12,2	+ 0,2	21,1	19,5	+ 1,6
15 »	10,9	11,0	— 0,1	14,2	13,7	+ 0,5	18,7	21,1	— 2,4
1er juillet . . .	11,8	11,9	— 0,1	14,7	14,8	— 0,1	20,7	22,1	— 1,4
15 » . . .	12,3	12,6	— 0,3	15,7	15,7	0	25,5	22,4	+ 3,1
1er août	13,1	13,2	— 0,1	16,6	16,1	+ 0,5	24,5	22,1	+ 2,4
15 »	13,6	13,6	0	16,5	16,1	+ 0,4	18,2	21,1	— 2,9
1er septembre .	13,9	13,8	+ 0,1	16,0	15,7	+ 0,3	20,2	19,5	+ 0,7
15 » .	13,9	13,8	+ 0,1	14,7	14,8	— 0,1	17,8	17,4	+ 0,4
1er octobre . .	13,7	13,6	+ 0,1	13,1	13,7	— 0,6	15,4	15,1	+ 0,3
15 » . .	13,3	13,2	+ 0,1	11,4	12,2	— 0,8	9,4	12,5	— 3,1
1er novembre .	12,5	12,6	— 0,1	10,2	10,6	— 0,4	7,1	9,9	— 2,8
15 » .	12,0	11,9	+ 0,1	8,6	9,0	— 0,4	7,2	7,6	— 0,4
1er décembre .	11,1	11,0	+ 0,1	6,6	7,4	— 0,8	5,0	5,5	— 0,5
15 » .	10,4	10,2	+ 0,2	6,4	6,0	+ 0,4	6,4	3,9	+ 2,5
En moyenne . .	10,63	10,63	0,13	9,8	9,8	0,38	12,5	12,5	1,47

$12°,5$ étant la température annuelle moyenne de l'air atmosphérique et présentant l'ordonnée de la sinusoïde au 8 avril ($0°$) et au 8 octobre ($180°$) (*).

Afin que le lecteur puisse se rendre compte clairement de la concordance ou de la divergence entre les données réelles et ces formules de péréquation, nous développons ces formules et nous mettons sous les yeux du lecteur les séries réelles et les séries de péréquation représentées numériquement et graphiquement.

On voit ainsi que même pour l'air atmosphérique, il existe encore une notable concordance entre les chiffres réels et les ordonnées de la sinusoïde représentée par notre formule [106] ; en moyenne, l'écart entre les chiffres réels et ceux obtenus par péréquation est ici d'un degré et demi ($1°,47$). Mais la concordance est beaucoup plus stricte pour le sol, surtout pour le sol à 3 mètres de profondeur : l'écart moyen entre les chiffres réels et les chiffres théoriques (les ordonnées de la sinusoïde) n'est que d'un tiers de degré ($0°,38$) pour le sol à 1^m de profondeur et de 1/8 de degré ($0°,13$) seulement pour le sol à 3^m de profondeur. Relativement aux dimensions des phénomènes observés $\left(\frac{\Sigma\,(v - y)}{\Sigma\, v}\right.$ ou $\frac{\Sigma\,(v - y)}{\Sigma\, y}$ ou encore $\frac{\Sigma\,(v - y) : 24}{\Sigma\, v : 24}$, ce qui revient dans la réalité au même$\Big)$, l'écart entre les chiffres réels et ceux obtenus par la péréquation est :

pour l'air atmosphérique	= 0,117
pour le sol, à 1^m de profondeur	= 0,039
pour le sol, à 3^m de profondeur	= 0,0125

Pour l'air atmosphérique, la courbe réelle s'écarte ainsi de 12 % de la sinusoïde ; pour le sol à 1^m de profondeur, l'écart est à peine de 1/25 (ou 4 %) ; pour le sol à 3^m de profondeur, la courbe réelle coïncide avec la sinusoïde à 1/80 près (ou 1 ¼ %). Et cela malgré le fait que la série réelle présente les moyennes de cinq années seulement.

Naturellement, par rapport à la seule partie *variable* des phénomènes observés $\left(\text{c'est-à-dire } \frac{\Sigma\,(v - y)}{\Sigma\,(b \sin x)} \text{ ou } \frac{\Sigma\,(v - y)}{\Sigma\, d}\right)$, l'écart entre

(*) Si on met le point 0 au 15 avril, la formule de péréquation est, pour l'air atmosphérique, comme suit :

$$y = 12°,5 + 9°,9 \sin x$$

les chiffres réels et ceux de la sinusoïde se montre beaucoup plus considérable (Cf. plus haut, § 26, première note en bas de la page) : pour l'air atmosphérique cet écart relatif s'élève à 0,24 ; il est de 0,09 pour le sol à 1^m de profondeur et n'est que de 0,06 pour le sol à 3^m de profondeur (*). En compliquant un peu la formule de péréquation, on pourrait, pour le sol à 3^m de profondeur, obtenir même une coïncidence presque complète entre la formule et la réalité (**) ; mais n'insistons pas.

Les observations relatives à la péréquation des trois périodicités par le procédé de la sinusoïde peuvent être résumées par le petit tableau suivant :

	Formule de la sinusoïde	Situation du point 0°	Ecart moyen entre les chiffres réels et la sinusoïde		
			en degrés Celsius	relativement aux grandeurs observées	relativement à la partie variable des grandenrs observées
Air atmosphèrique	$12°,5 + 9°,7 \sin x$	8(15)/IV	1°,47	0,117	0,24
Sol, profondeur 1^m	$9°,8 + 6°,3 \sin x$	8 /V	0°,38	0,039	0,09
Sol, profondeur 3^m	$10°,6 + 3°,2 \sin x$	8 /VI	0°,13	0,013	0,06

(*) Si pour l'air atmosphérique on pose le point 0 au 15 avril (au lieu du 8), les écarts sont respectivement de 1°,51 (au lieu de 1°,47), de 0,121 (au lieu de 0,117) et de 0,241 (au lieu de 0,237) ; les différences, on le voit, sont négligeables.

(**) D'après notre dernier tableau (ou d'après le graphique qui sert à l'illustrer), on peut en effet remarquer que, pour le sol à 3^m de profondeur, notre courbe théorique est pendant la phase de hausse (avril-août) continuellement un peu au-dessus des chiffres réels et qu'elle se trouve systématiquement un peu au-dessous de la réalité pendant la phase de baisse (septembre-février), ce qui signifie que la formule pèche par un léger excès (hausse et baisse trop forte). Complétant dans ce sens notre formule [104] d'après un procédé bien connu, nous obtenons pour le sol à 3^m de profondeur la formule :

$$y' = 10°,63 + 3\,,2 \sin x - 0°,184 \cos x$$

qui donne des chiffres théoriques coïncidant remarquablement avec les chiffres réels (pour 7 termes, sur les 24, l'écart entre les chiffres réels et la formule est $= 0$, pour 14 termes, il est $= \pm 0°,1$ et seulement pour 3 l'écart est $= \pm 0°,2$).

Ce petit tableau nous montre clairement dans quelle mesure les variations considérées se rapprochent de la sinusoïde ; nous voyons en particulier que relativement aux grandeurs observées l'écart entre les chiffres réels et la sinusoïde est pour le sol à 1 mètre de profondeur trois fois moindre que pour l'air atmosphérique, et pour le sol à 3 mètres trois fois moindre que pour le sol à 1 mètre (0,013 : 0,039 : 0,117)

Il nous montre aussi que la périodicité de la température du sol à 1^m de profondeur est de trois semaines à un mois en retard sur celle de l'air atmosphérique et que celle du sol à 3^m de profondeur est d'un mois en retard sur celle du sol à 1^m de profondeur (situation du point 0^o : le 8 (15) /IV, le 8 /V et le 8 /VI).

Il nous montre encore que la température annuelle moyenne est pour le sol à 1^m de profondeur à la fois moins élevée que celle de l'air atmosphérique et sensiblement moindre que celle du sol même à 3^m de profondeur.

Il permet enfin de comparer l'intensité de l'écartement moyennal de ces trois séries de variations périodiques : il montre en effet que cet écartement moyennal est absolument parlant (en degrés Celsius) pour le sol à 1^m deux fois et pour l'air atmosphérique trois fois plus fort que pour le sol à 3^m de profondeur (le paramètre de la sinusoïde b étant égal respectivement à $9^o,7$, à $6^o,3$ et à $3^o,2$).

Pour être beaucoup moins détaillé que notre tableau du § 70, ce petit tableau n'en est donc pas moins assez instructif.

En outre, pour autant que les données réelles s'écartent très peu de la sinusoïde (comme dans le cas de la température du sol), ce petit écart comporte implicitement aussi l'indication que la série considérée présente à un haut degré les divers caractères de la sinusoïde qui ne sont pas mentionnés sur ce petit tableau, tels que la monocéphalie, la gradualité, la symétrie, etc. Mais plus cet écart est important, et moins nous sommes renseignés par une telle méthode sur ces autres caractères des périodicités considérées. Si l'écart entre les chiffres réels et la sinusoïde est considérable, la formule de la péréquation par la sinusoïde devenant pour une large part fictive, la comparaison des données qui figurent même explicitement sur un pareil tableau (comme nous venons de le faire ici) peut facilement aboutir à des conclusions erronées. C'est encore la série des indices et des coefficients qui figurent sur notre tableau § 70 qui se montre d'une application plus générale, plus sûre et plus précise.

§ 75. — Après l'analyse relativement détaillée de l'exemple précédent, nous serons bref pour les autres.

Prenons maintenant un exemple réel emprunté au domaine *démographique*. Examinons notamment la *mortalité infantile* selon les mois dans trois pays différents : en Suisse, en France et dans l'U. R. S. S.

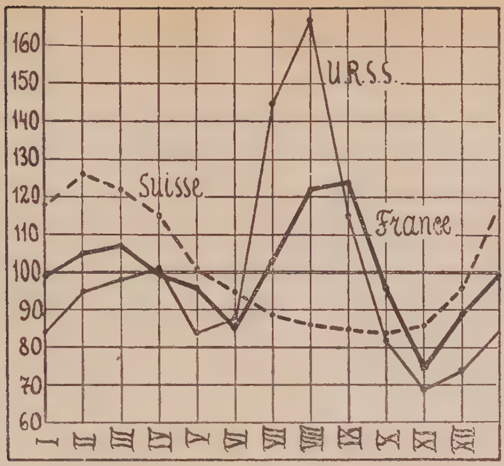
On sait que la mortalité en général (comme d'ailleurs tous les phénomènes démographiques) et la mortalité infantile en particulier accusent une périodicité annuelle bien stricte. De nos jours, la mortalité infantile accuse dans les divers pays de notre hémisphère un premier maximum à la fin de l'hiver ou tout au début du printemps ; dans le temps, un second maximum de la mortalité infantile qui souvent dépassait même le premier, pouvait être constaté partout pendant les mois les plus chauds de l'année. La remarquable baisse de la mortalité infantile que l'on constate dans les pays de la civilisation occidentale depuis un demi-siècle (*) a porté tout particulièrement sur la mortalité causée par les chaleurs d'été. De cette façon le maximum estival allait s'atténuant à mesure que baissait la mortalité infantile ; il devenait de plus en plus secondaire, et dans les pays où cette mortalité est aujourd'hui particulièrement faible, il a fini par disparaître complètement. Or, les pays que nous venons de nommer marquent des stades très différents dans la baisse de la mortalité infantile : l'U. R. S. S. a encore une mortalité infantile extrêmement élevée, un cinquième (20,5 p. 100 en 1926-1927) de tous les enfants nés vivants dans ce pays mourant avant d'avoir atteint une année de vie ; cette mortalité est déjà beaucoup plus faible en France (8,9 p. 100 de tous les nés vivants en 1926-1930); la Suisse est aujourd'hui un des pays du monde entier où la mortalité infantile est la plus faible (5,4 décès d'enfants de moins d'un an pour 100 nés vivants en 1926-1930). Sur un nombre égal de nouveau-nés, là où 100 enfants meurent en Suisse, 165 en meurent en France et 380 dans l'U. R. S. S. Il serait dans ces conditions intéressant d'examiner et de comparer les séries des variations mensuelles de la mortalité infantile dans ces trois pays d'après les méthodes et le schéma exposés plus haut. (Pour les variations mensuelles, voy. le tableau II de l'Annexe II et le graphique XII).

(*) En Suisse, sur 100 enfants nés vivants moururent dans la première année de la vie 18,8 en 1876-80 et seulement 4,9 en 1931, soit une proportion quatre fois moindre.

Caractères de la périodicité annuelle de la mortalité infantile en Suisse (1926-1930), en France (1926-1928) et dans l'U. R. S. S. (1926-1927).

	Suisse	France	U.R.S.S.
Nombre quotidien moyen de décès (v_m)	10,36	187,1	2439
Partie de l'année ayant $v > v_m$	0,42	0,42	0,33
Partie de l'année ayant $v < v_m$	0,58	0,58	0,67
Nombre des sommets	1	2	2
Maximum principal :			
valeur (v_{max})	13,1	231,5	4063
moment de l'année (mois) .	II	IX	VIII
Minimum principal :			
valeur (v_{min})	8,7	139,7	1676
moment de l'année	X	XI	XI
Maximum secondaire :			
valeur ($v_{max'}$)	—	199,5	2474
moment de l'année	—	III	IV
Minimum secondaire :			
valeur ($v_{min'}$)	—	159,8	2041
moment de l'année	—	VI	V
Amplitude des variations :			
α	4.4	91,8	2387
$\frac{\alpha}{v_m}$	0,42	0,49	0,98
$\frac{\alpha}{2\sigma}$	1,39	1,84	1,75

	Suisse	France	U.R.S.S.
Écartement moyennal relatif :			
$\frac{d_m}{v_m}$	0,135	0,10	0,21
$\frac{\sigma}{v_m}$	0,15	0,13	0,28
Écartement successif relatif :			
$\frac{e_m}{v_m}$	0,07	0,12	0,19
$\frac{\varsigma}{v_m}$	0,09	0,14	0,26
Concentration et nivellement :			
coefficient de concentration k_{12}	0,24	0,46	0,40
» » » c_{12}	0,09	0,22	0,22
» » nivellement v_{12}	0,81	0,59	0,59
Coefficients de :			
précipitation (p_{12})	0,20	—	—
gradualité (g_{12}) :	0,62	—	—
Coefficients de symétrie (s) :			
des écarts moyennaux (s_d) . .	0,94(I/III)	0.90(I IX)	0,69(I/VIII)
des écarts successifs (s_e) . . .	0,89(I/III)	0,94(I/IX)	0,75(VIII)
des écarts centraux (s_η) . . .	0,94(II)	0,94(I/IX)	0,96(VIII)



Graphique XII

Variations mensuelles de la mortalité infantile en % % de la moyenne de l'année.

Les nombres *absolus* des décès d'enfants de moins d'un an survenus dans ces trois pays ne peuvent naturellement pas être comparés tels quels étant donné l'énorme différence entre les chiffres des naissances de ces pays. Il n'est cependant pas sans intérêt de comparer les différences qui existent entre ces trois pays pour différentes espèces de chiffres absolus. Nous pouvons ainsi constater que la moyenne (v_m), le minimum principal (v_{min}), le maximum principal (v_{max}) et l'amplitude des variations (α) de l'année se trouvent pour les trois pays dans les rapports que voici :

	Suisse	France	U. R. S. S.		Suisse	France	U. R. S. S.
v_m	10,36 :	187,1 :	2439	=	1 :	18 :	235
v_{min}	8,7 :	139,7 :	1676	=	1 :	16 :	192
v_{max}	13,1 :	231,5 :	4063	=	1 :	18 :	310
α	4,4 :	91,8 :	2387	=	1 :	21 :	543

Nous voyons ainsi que ceux de ces trois pays qui ont une mortalité infantile plus forte ont une amplitude de variations saisonnières plus forte encore. On voit aussi qu'au point de vue de leur minima de mortalité infantile, ces pays diffèrent moins que sous le rapport de leur

mortalité infantile moyenne, et que leur maxima diffèrent au contraire plus que les moyennes annuelles. La grande différence des maxima fait ainsi monter la moyenne annuelle le plus fortement dans les pays ayant un maximum particulièrement élevé, et il est dès lors compréhensible aussi que l'amplitude des variations varie d'un pays à l'autre plus fortement que la moyenne annuelle. Les maxima saisonniers élevés paraissent ainsi un des *facteurs* essentiels de la forte mortalité infantile moyenne de l'année. (Nous l'observons au sujet des trois pays considérés, mais il y aurait lieu naturellement de le vérifier aussi pour les autres pays et, le cas échéant, d'établir sous ce rapport un classement des divers pays).

Ces caractères se trouvent d'ailleurs précisés et complétés par les autres indices qui figurent sur notre tableau. Signalons les plus caractéristiques d'entre eux.

La courbe de la mortalité infantile de la Suisse est monocéphale, celles de la France et de l'U. R. S. S. sont bicéphales. Le minimum principal tombe dans les trois pays en automne (octobre-novembre) ; mais le maximum principal de la France et de l'U. R. S. S., qui est un maximum estival (septembre en France et août dans l'Union Soviétique), n'existe plus du tout en Suisse, le maximum principal de la Suisse (février) correspondant plus ou moins au maximum secondaire de la France (mars) et de l'U. R. S. S. (avril).

Entre la mortalité infantile de la Suisse et celle de l'U. R. S. S., la différence est énorme sous tous les rapports. Au point de vue de *l'amplitude* des variations, de leur écartement *successif* et aussi de la *symétrie* des écarts moyennaux, la mortalité infantile de la France occupe une place intermédiaire entre celle de la Suisse, d'un côté, et de l'U. R. S. S., de l'autre. En effet, même relativement à la mortalité infantile moyenne de l'année (moyenne qui est déjà très faible en Suisse, qui est plus importante en France et qui est excessivement élevée dans l'U. R. S. S.), l'amplitude des variations $\left(\frac{\alpha}{v_m}\right)$ est encore la plus faible en Suisse, un peu plus élevée en France et presque deux fois et demie plus forte dans l'Union Soviétique (respectivement : 0,42, 0,49 et 0,98). L'écartement successif est en Suisse, même relativement $\left(\frac{e_m}{v_m} \text{ ou } \frac{\varsigma}{v_m}\right)$ beaucoup plus faible qu'en France et presque trois fois moindre que dans l'U. R. S. S. ($\frac{e_m}{v_m}$ étant respectivement 0,07, 0,12 et 0,19) ; cette différence entre les trois pays est

liée à la monocéphalie de la courbe suisse et à la bicéphalie des courbes française et soviétique, aggravée encore dans l'U. R. S. S. par la hauteur extraordinaire du maximum estival.

Cependant, l'écartement *moyennal* moyen, pris lui aussi relativement à la mortalité infantile annuelle moyenne $\left(\frac{d_m}{v_m} \text{ ou } \frac{\sigma_m}{v_m}\right)$ est en France légèrement inférieur à celui observé en Suisse (respectivement 0,13 et 0,15 en Suisse et 0,10 et 0,13 en France). C'est qu'en France les variations mensuelles de la mortalité infantile ont un degré de *concentration* beaucoup plus élevé (ou un nivellement beaucoup moindre) qu'en Suisse. En effet, le coefficient de concentration k_{12} est égal à 0,24 en Suisse et à 0,46 en France, le coefficient c_{12} est de 0,09 en Suisse et de 0,22 en France ; le coefficient de nivellement ν_{12} est de 0,81 en Suisse et de 0,59 en France. Sous le rapport de la concentration (ou du nivellement) des variations mensuelles de la mortalité infantile, la France ressemble bien à l'U. R. S. S. qui a exactement les mêmes coefficients c_{12} et ν_{12}.

La courbe suisse étant la seule monocéphalique, nos coefficients de précipitation (p_{12}) et de gradation (g_{12}) ne peuvent être calculés ni pour la France ni pour l'U. R. S. S. La courbe suisse, qui d'après notre graphique XII est manifestement moins précipitée que les deux autres, a un coefficient de précipitation assez faible ($p_{12} = 0,20$) et un coefficient de gradualité relativement fort ($g_{12} = 0,62$). Cependant, la courbe suisse même est encore plus précipitée (ou moins graduelle) que les courbes de la température examinées au § 70, plus précipitée même que celle de la température de l'air atmosphérique, la plus précipitée de toutes les trois (qui avait $p_{12} = 0,15$ et $g_{12} = 0,70$).

La différence entre les trois pays envisagés est moins forte au point de vue de la symétrie des variations. En effet, toute série périodique conserve une importante symétrie, forte surtout pour les écarts centraux, puisqu'à la fin de la série on revient exactement à son point d'origine. Grâce à ce caractère des séries périodiques, la symétrie des écarts centraux se montre particulièrement élevée là où le maximum (ou le minimum) principal s'écarte très fortement du reste de la série : si l'on prend ce maximum (ou ce minimum) comme centre de symétrie des écarts centraux, leur degré de symétrie se montrera forcément très élevé. C'est ce que nous voyons aussi dans le cas particulier de l'U. R. S. S.: en prenant comme centre de symétrie des écarts centraux le terme maximum de la série (celui du mois d'août), on obtient le coefficient $s_\eta = 0,96$ (contre 0,94 pour la Suisse et la

France). Mais le coefficient de symétrie des écarts successifs (s_e) n'est pour l'U. R. S. S. que de 0,75 et celui des écarts moyennaux (s_d) n'est même que 0,69, coefficients assez bas pour une série périodique et bien inférieurs à ceux de la Suisse et de la France (qui varient entre 0,89 et 0,94).

Il ne saurait être question de résumer les périodicités annuelles de la mortalité infantile en France et dans l'U. R. S. S. en les représentant par des *sinusoïdes*, puisque ce sont des périodicités nettement bicéphales qui, de plus, diffèrent de la sinusoïde sous bien d'autres rapports encore. La périodicité annuelle de la mortalité infantile en Suisse se rapproche beaucoup plus d'une sinusoïde : elle est monocéphale, elle est moins concentrée, moins précipitée et plus symétrique que la périodicité du même phénomène dans les deux autres pays. En outre, le rapport $\frac{\alpha}{2\,\sigma,}$ qui pour la sinusoïde est, d'après [13], égal à 1,41, est pour la mortalité infantile en Suisse égal presque au même chiffre : 1,39. Pourtant, ici encore la série est loin de présenter une symétrie parfaite, elle est plus loin encore d'être tétra-symétrique ; la partie de la période ayant $v > v_m$ n'est pas égale à $^1/_2$, de même que la distance qui sépare le maximum (février) du minimum (octobre) ou vice versa n'est pas égale à la moitié de la période : en faisant la péréquation par la sinusoïde, on déplacerait donc arbitrairement le moment du maximum ou celui du minimum ou tous les deux ; le coefficient de précipitation est supérieur à ceux que nous avons trouvés plus haut non seulement pour la température du sol, mais même pour la température de l'air.

Nous avons pourtant calculé la sinusoïde de péréquation de la mortalité infantile en Suisse. En mettant le point 0° au 1^er^ décembre (soit en mettant le maximum, correspondant à 90°, au 1^er^ mars, qui constitue notre centre de symétrie des écarts moyennaux et successifs) ou, en d'autres termes, en faisant correspondre le chiffre de décembre à 15°, celui de janvier à 45°, celui de février à 75° et ainsi de suite, on trouve pour la sinusoïde cherchée la formule suivante :

$$y = 10{,}36 + 2{,}09 \sin x \qquad [107]$$

En mettant les ordonnées de cette sinusoïde en regard des chiffres réels de la mortalité infantile en Suisse, nous obtenons le petit tableau que voici :

Nombre quotidien moyen de décès d'enfants de moins d'un an en Suisse (1926-1930) *selon les mois.*

Mois	Chiffres réels (v)	Sinusoïde (y)	v-y	Mois	Chiffres réels (v)	Sinusoïde (y)	v-y
Janvier	12,2	11,8	+ 0,4	Juillet	9,2	8,9	+ 0,3
Février	13,1	12,4	+ 0,7	Août	8,9	8,3	+ 0,6
Mars	12,6	12,4	+ 0,2	Septembre . . .	8,8	8,3	+ 0,5
Avril	11,9	11,8	+ 0,1	Octobre	8,7	8,9	— 0,2
Mai	10,5	10,9	— 0,4	Novembre . . .	8,9	9,8	— 0,9
Juin	9,8	9,8	0	Décembre . . .	9,9	10,9	— 1,0

Les différences entre les chiffres réels et ceux de la sinusoïde ne paraissent pas très grandes. En moyenne $\left(\text{c'est-à-dire : } \frac{\Sigma\,(v - y)}{\Sigma\, v}\right)$, elles forment 0,04 de la mortalité réelle. Cependant, par rapport à la seule partie variable de cette mortalité $\left(\text{soit } \frac{\Sigma\,(v - y)}{\Sigma\, d}\right)$, la différence entre les variations réelles et les ordonnées de la sinusoïde est, pour la Suisse aussi, vraiment considérable, étant égale à 0,31 ; elle est sous ce rapport beaucoup plus grande que ce que nous avons vu plus haut (§ 74) même pour les variations de la température de l'air atmosphérique soumise à des observations bi-mensuelles. Et encore faudrait-il tenir compte des autres altérations de la courbe réelle (déplacement du maximum et du minimum, leur atténuation, etc.) que nous avons déjà mentionnées et qui sont à peine admissibles quand il s'agit d'une série typique.

§ 76. — Prenons maintenant un phénomène d'ordre économique et social dans le sens étroit du terme : le *chômage*. Examinons notamment les fluctuations mensuelles du pour-cent de chômeurs au Royaume-Uni (parmi les ouvriers assurés) et au Danemark (parmi les syndiqués) dans l'industrie du bâtiment en général et, plus spécialement, chez les peintres en bâtiment. (Pour les chiffres mensuels, voyez le tableau III de notre Annexe II). Quels sont les principaux caractères de ces périodicités ?

Caractères de la périodicité annuelle du taux de chômage dans l'industrie du bâtiment et spécialement chez les peintres en bâtiment au Royaume-Uni (1913-1927) et au Danemark (1910-1926).

	Royaume-Uni		Danemark	
	Bâtiment	Peintres	Bâtiment	Peintres
Taux moyen de l'année (v_m)	7,5	10,4	16,7	20,9
Partie de l'année ayant :				
$v > v_m$	0,42	0.42	0,42	0,42
$v < v_m$	0,58	0,58	0,58	0,58
Nombre des sommets	1	1	2	2
Maximum principal :				
valeur (v_{max})	9.5	20,5	40,2	55,3
moment *	I	I	I	I
Minimum principal :				
valeur (v_{min})	5,9	2,8	5,2	1,0
moment *	VI	V	VII	IV
Maximum secondaire :				
valeur	—	—	7,0	3,5
moment *	—	—	VIII	VI
Minimum secondaire :				
valeur	—	—	6,6	3,2
moment *	—	—	IX	VII
Amplitude des variations :				
α	3,6	17,7	35,0	54,3
$\frac{\alpha}{v_m}$	0,48	1,70	2,10	2,60
$\frac{\alpha}{2\sigma}$	1,47	1,42	1,36	1,32
Écartement moyennal :				
absolu d_m	1,1	5,6	11,5	18,5
absolu σ	1,2	6,2	12,9	20,6
relatif $\frac{d_m}{v_m}$	0,15	0,54	0,69	0,88
relatif $\frac{\sigma}{v_m}$	0,16	0,60	0,77	0,99
Écartement successif :				
absolu e_m	0,62	2,95	5,9	9,1
absolu ς	0,68	3,64	8,0	13,2
relatif $\frac{e_m}{v_m}$	0,08	0,28	0,35	0,44
relatif $\frac{\varsigma}{v_m}$	0,09	0,35	0,48	0,63
Coefficients de :				
concentration (k_{12})	0,25	0,25	0,22	0,22
» (c_{12})	0,07	0,08	0,08	0,08
nivellement (v_{12})	0,85	0,83	0,82	0,82
précipitation (p_{12})	0,07	0,16	0,25**	0,31**
gradualité (g_{12})	0,85	0,68	0,55*	0,48**
Coefficients de symétrie :				
s_d valeur	0,85	0,76	0,84	0,96
s_d centre	15/II	15.I	15/II	15/I
s_e valeur	0,91	0,82	0,96	0,89
s_e centre*	I	XII	I	I
s_n valeur	0,994	0,986	0,998	0,977
s_n centre*	I	XII	I	I

* Il s'agit de la fin du mois considéré.

** Nous avons calculé ces coefficients aussi pour le Danemark, les maxima secondaires formant des fluctuations d'importance minime (dépassant les termes suivants de 0,4 et de 0,3).

Comparant le taux du chômage dans les branches considérées tel qu'il a été observé dans les deux pays à l'époque envisagée, nous pouvons faire ainsi les constatations suivantes :

Le taux annuel moyen du chômage (v_m) dans les branches en question a été pour le Danemark plus du double de celui observé au Royaume-Uni (pour l'ensemble du bâtiment, 16,7 : 7,5 = 2,23 ; pour les peintres 20,9 : 10,4 = 2,01). Les fluctuations mensuelles du taux du chômage étaient encore beaucoup plus différentes d'un pays à l'autre. L'amplitude de ces fluctuations (α), l'écart moyennal moyen, tant arithmétique (d_m) que quadratique (σ), ainsi que l'écart successif moyen (e_m ou ς) ont été chez les peintres au Danemark plus du triple de ceux observés pour la même profession au Royaume-Uni ; pour l'ensemble du bâtiment, l'amplitude des fluctuations mensuelles du chômage (α), l'écartement moyennal moyen (d_m et σ) et l'écartement successif moyen (e_m et ς) ont été même 10 à 12 fois plus forts au Danemark qu'au Royaume-Uni. Même relativement au taux annuel moyen, bien plus élevé au Danemark qu'au Royaume-Uni, les fluctuations mensuelles du chômage des branches considérées se montrent beaucoup plus considérables au Danemark. Pour les peintres, $\frac{\alpha}{v_m}$, $\frac{\sigma}{v_m}$ et $\frac{\varsigma}{v_m}$ sont au Danemark de 50 à 80 p. 100 plus élevés qu'au Royaume-Uni ; pour l'ensemble du bâtiment, ces mêmes grandeurs relatives sont dans le premier pays 4 à 5 fois plus élevées que dans le second.

Les coefficients de précipitation des fluctuations mensuelles (p_{12}) sont de même beaucoup plus élevés (ou les coefficients de gradualité beaucoup plus faibles) au Danemark qu'au Royaume-Uni, et cela aussi bien pour les peintres que pour l'ensemble des travailleurs du bâtiment. On constate, en outre, au Danemark un certain maximum secondaire en été qui n'existe pas au Royaume-Uni (ce second maximum est pourtant au Danermak aussi d'une importance presque négligeable : il ne dépasse le taux du mois suivant que de 0,4 — en août — pour le bâtiment et de 0,3 — en juin — pour les peintres, il demeure une petite fraction du taux annuel moyen et appartient aux taux mensuels les plus bas).

Et malgré ces énormes différences entre les dimensions annuelles, comme entre les fluctuations saisonnières, du taux de chômage dans l'industrie du bâtiment des deux pays, nous pouvons constater que bien des caractères de ce chômage saisonnier se retrouvent aussi bien au Danemark qu'au Royaume-Uni.

Nous pouvons ainsi constater que dans les deux pays le maximum de chômage se place à la fin du mois de janvier, et cela aussi bien pour l'ensemble de l'industrie du bâtiment que pour la profession des peintres. — Pour les deux pays, nous pouvons constater que, dans l'ensemble du bâtiment, le chômage baisse à partir du mois de février jusqu'aux mois juin-juillet pour remonter dans la suite lentement jusqu'à la fin d'octobre et brusquement depuis la fin de novembre. — Pour les deux pays et pour les deux catégories pofessionnelles, nous constatons que le chômage dépasse la moyenne annuelle pendant moins de temps (5 mois) qu'il ne reste au-dessous d'elle (7 mois) et que par conséquent les écarts en plus de la moyenne sont plus forts que les écarts en moins. — Nous constatons encore que malgré la grande différence qui existe entre les deux pays au point de vue de la précipitation (ou de la gradualité) des variations mensuelles du chômage, le degré de concentration de ces variations reste dans les deux pays très faible et s'exprime pour les deux pays et pour les deux catégories professionnelles par un chiffre presque constant ($k_{12} = 0{,}22$ à $0{,}25$; $c_{12} = 0{,}07$ à $0{,}08$; $\nu_{12} = 0{,}82$ à $0{,}85$). — Dans les deux pays, les fluctuations mensuelles du chômage présentent des séries qui sans être parfaitement symétriques, le sont pourtant à un haut degré.

On constate de même que dans les deux pays les peintres accusent non seulement un taux de chômage annuel moyen (v_m) supérieur à celui de l'ensemble du bâtiment, mais encore et surtout que les fluctuations saisonnières du chômage sont chez eux beaucoup plus fortes. Au Danemark, où le bâtiment accuse en général des fluctuations saisonnières formidables, l'amplitude des variations et les écarts moyens, tant moyennaux que successifs, sont pour les peintres de 60 p. 100 environ plus forts encore que pour l'ensemble du bâtiment ; au Royaume-Uni, où l'ensemble du bâtiment accuse des fluctuations saisonnières beaucoup moindres qu'au Danemark, les fluctuations saisonnières du chômage (α, d_m, σ, e_m et ς) sont pour les peintres cinq fois plus fortes que pour le bâtiment en général. Même relativement au taux annuel moyen du chômage (v_m) qui pour les peintres est plus élevé que pour l'ensemble des travailleurs du bâtiment, les fluctuations saisonnières se montrent bien plus fortes pour les peintres que pour le bâtiment en général: de quelques 30 p. 100 au Danemark et de 3 à 4 fois au Royaume-Uni. On constate encore que dans les deux pays, la baisse saisonnière du chômage finit pour les peintres plus tôt (tout en tombant plus bas) que pour l'ensemble des travailleurs du bâtiment (minimum absolu fin mai au Royaume-Uni et fin avril au Da-

nemark). Aussi, peut-on constater pour les deux pays que la précipitation des variations saisonnières du chômage est pour les peintres bien plus forte que pour le bâtiment en général ($p_{12} = 0{,}31$, pour les peintres et 0,25 pour l'ensemble du bâtiment au Danemark ; au Royaume-Uni, nous trouvons respectivement 0,16 contre 0,07). Comme le montre notre dernier tableau, on peut dire que les fluctuations *relatives* du chômage saisonnier des peintres du Royaume-Uni ressemble plus à celles observées chez les peintres du Danemark qu'à celles constatées chez les autres travailleurs du bâtiment dans leur propre pays (*).

§ 77. — Prenons encore un phénomène du domaine techniquement économique, et nous nous arrêterons. Nous envisagerons la consommation journalière moyenne de *gaz* d'éclairage à Genève (servant presaque exclusivement au chauffage).

La consommation de gaz à Genève, comme dans bien d'autres villes, subit des variations déterminées non seulement par la tendance générale vers la hausse, par la conjoncture économique et par des circonstances fortuites, mais encore par une triple périodicité, fortement prononcée et strictement contrôlée :

1° une périodicité *annuelle* caractérisée par une forte baisse en hiver (à cause d'une plus grande consommation de charbon pour les cuisines), avec le minimum principal en janvier, et par une forte

---

(*) Etant donné la grande précipitation des variations mensuelles constatées pour le Danemark et pour les peintres du Royaume-Uni, étant donné encore l'absence presque complète de tétra-symétrie dans ces variations ainsi que la valeur élevée de $\frac{\alpha}{v_m}$, on ne saurait représenter ces fluctuations par une simple sinusoïde. Pour le chômage dans l'ensemble du bâtiment au Royaume-Uni (qui a une allure beaucoup plus graduelle, ondulatoire, et où $\frac{\alpha}{v_m}$ a une valeur relativement faible, la péréquation par une sinusoïde ayant pour formule : $y = 7{,}5 + 1{,}69 \sin x$ donne une série de chiffres qui s'écarte assez peu de la série typique réelle $\left(\text{en effet } \frac{\Sigma |v - y|}{\Sigma v} = 3 \text{ p. } 100\right)$; cependant même ci, l'écart entre la péréquation et la réalité envisagé par rapport à la seule partie variable du chômage demeure considérable $\left(\frac{\Sigma |v - y|}{\Sigma d} = 20 \text{ p. } 100\right)$, sans parler même des autres inconvénients d'une telle péréquation. Pour cette péréquation, voyez *Les Fluctuations saisonnières du chômage dans l'industrie du bâtiment*, p. 36).

hausse au début de l'été, avec son maximum en juin, suivie d'une seconde baisse en juillet-août (nombreux départs en villégiature) et un nouveau maximum (à peu près égal au premier) en septembre ;

2º une périodicité *hebdomadaire*, accusant elle aussi de fortes variations, selon les jours de la semaine, et

3º une périodicité *diurne* accusant des variations plus fortes encore, selon les heures de la journée.

Ces fluctuations créent des problèmes économiques et techniques importants, car les installations techniques de l'usine à gaz ainsi que son personnel et, par suite, aussi les frais généraux doivent être modelés sur les besoins des moments de la consommation maxima, ce qui rehausse dans des proportions considérables le coût de la production et les prix du gaz. Ici (puisqu'il s'agit principalement de la démonstration de la méthode), nous nous arrêterons surtout sur la périodicité *hebdomadaire* de la consommation du gaz présentant une série périodique de 7 termes.

Nous nous demanderons d'abord si les variations selon les jours de la semaine sont les mêmes pendant la baisse d'hiver et la hausse du début de l'été et nous comparerons ensuite ces périodicités hebdomadaires (de la baisse hivernale et de la hausse printanière) avec la périodicité annuelle du même phénomène. Pour établir les deux séries hebdomadaires typiques, nous avons pris la consommation moyenne, par jour de la semaine, d'une part, pendant les huit semaines allant du 3 janvier au 27 février (inclusivement) et, de l'autre, pendant les huit semaines allant du 2 mai au 26 juin (incl.), le tout selon les données de 1927 et en commençant dans les deux cas par un lundi. Pour la périodicité annuelle, nous avons simplement admis comme typiques les données de 1927. Nous avons obtenu ainsi le tableau IV de notre Annexe II. — Selon notre schéma, les caractères de ces variations périodiques peuvent être représentés de la façon suivante :

Caractères des périodicités hebdomadaires et annuelle de la consommation du gaz d'éclairage à Genève (1927).

(Consommation journalière moyenne en milliers de m³).

	Périodicité hebdomadaire en janvier-février	Périodicité hebdomadaire en mai-juin	Périodicité annuelle
Consommation journalière moyenne de la période (v_m)	41,7	53,6	47,8
Partie de la période ayant :			
$v > v_m$	0,29	0,29	0,50
$v < v_m$	0,71	0,71	0,50
Nombre des sommets	2	2	2
Maximum principal :			
valeur (v_{max})	43.1	55,7	54,1
moment	samedi	samedi	IX
Minimum principal :			
valeur (v_{min})	38,1	48,9	40,9
moment	dimanche	dimanche	I
Maximum secondaire :			
valeur	42,7	55,4	54,0
moment	mardi (mercr.)	mercredi	VI
Minimum secondaire :			
valeur	42,0	54,6	49,2
moment	jeudi	vendredi	VIII
Amplitude des variations :			
α	5,05	6,75	13,2
$\frac{\alpha}{v_m}$	0,12	0,13	0,28
$\frac{\alpha}{2\sigma}$	1,56	1,52	1,53
Ecartement moyennal :			
absolu d_m	1,26	1,80	3,74
absolu σ	1,62	2,23	4,32

	Périodicité hebdomadaire en janvier-février	Périodicité hebdomadaire en mai-juin	Périodicité annuelle
Ecartement moyennal :			
relatif $\frac{d_m}{v_m}$	0,03	0,03	0,08
relatif $\frac{\sigma}{v_m}$	0,04	0,04	0,09
Ecartement successif :			
absolu e_m	1,65	2,17	3,00
absolu ς	2,33	2,99	3,60
relatif $\frac{e_m}{v_m}$	0,04	0,04	0,06
relatif $\frac{v_m}{\varsigma}$	0,06	0,06	0,07
Coefficients de :			
concentration (k_n)	0,40	0,35	0,32
» (c_n)	0,33	0,26	0,11
nivellement (v_n)	0,52	0,58	0,77
Coefficients de symétrie :			
s_d valeur	0,54	0,59	0,86
s_d centre	jeudi	jeudi	I, VIII
s_e valeur	0,69	0,60	0,60
s_e centre	mercredi-jeudi	mercredi-jeudi	I/VIII
s_η valeur	0,86	0,70	0,92
s_η centre	idem	idem	I/VIII

Les chiffres romains marquent les mois de l'année.

Comparant entre elles les périodicités *hebdomadaires*, celle observée pendant la baisse hivernale et celle constatée pendant la hausse de printemps, nous pouvons remarquer qu'en mai-juin les chiffres absolus sont de 30 p. 100 environ plus élevés qu'en janvier-février ; la consommation journalière moyenne (v_m) est à ces deux époques de l'année de 53.631 et de 41.688 m^3, soit dans le rapport de 129 à 100 ; le minimum principal de mai-juin (dimanche : 48.916) est encore bien supérieur au maximum principal de janvier-février (samedi : 43.108), et ainsi de suite. Malgré cela, la périodicité hebdomadaire garde pendant les deux époques de l'année les mêmes caractères essentiels. Aux deux époques, elle présente une courbe bicéphale, ayant son maximum principal le samedi et son minimum de beaucoup le plus profond le dimanche, et dans les deux cas le second maximum n'est que légèrement inférieur au maximum principal. — L'importance *relative* des fluctuations journalières est presque identique dans les deux cas : par rapport à la consommation journalière moyenne de la semaine, l'amplitude des variations $\left(\frac{\alpha}{v_m}\right)$ est dans les deux cas de 12 à 13 p. 100 (plus exactement 12,0 et 12,6) ; dans les deux cas, l'écart moyennal arithmétique moyen $\left(\frac{d_m}{v_m}\right)$ est de 3 p. 100 et l'écart moyennal quadratique moyen $\left(\frac{\sigma}{v_m}\right)$ est de 4 p. 100 ; dans les deux cas, l'écart successif arithmétique moyen a exactement la même valeur $\left(\frac{e_m}{v_m} = 4{,}0 \text{ p. } 100\right)$ de même que l'écart successif quadratique moyen $\left(\frac{\varsigma}{v_m} = 5{,}6 \text{ p. } 100\right)$. Dans les deux cas, il s'agit d'une courbe ayant une forte concentration, beaucoup plus forte que toutes les courbes examinées dans ce chapitre, la journée de dimanche accusant à elle seule presque la moitié de toute la somme des écarts moyennaux de la série). — Dans les deux cas la courbe accuse une symétrie relativement faible, plus faible que toutes les autres courbes de ce chapitre. La seule différence intéressante entre la périodicité hebdomadaire de mai-juin et celle de janvier-février (sans parler des dimensions absolues des fluctuations), c'est le fait que le second maximum comme le second minimum tombent à l'époque mai-juin un jour plus tard qu'en hiver (le second maximum est observé le mardi en janvier-février et le mercredi en mai-juin, le second minimum est respectivement le jeudi et le vendredi).

Si nous voulons établir un parallèle entre les périodicités hebdomadaires et la périodicité *annuelle* de la consommation du gaz, nous pouvons marquer les traits suivants :

Comme la périodicité hebdomadaire, la périodicité annuelle présente une courbe bicéphale ; ici encore, le second maximum (juin) atteint presque la hauteur du maximum principal (novembre) et, comme pour la périodicité hebdomadaire, le second minimum (août) est au contraire très loin de tomber au niveau du minimum principal. Sous les autres rapports, il existe cependant des différences essentielles entre les deux espèces de périodicités. D'un côté, la périodicité annuelle accuse des variations qui absolument et relativement sont beaucoup plus fortes que celles qui marquent la périodicité hebdomadaire. Ainsi, l'amplitude des variations $\left(\frac{\alpha}{v_m}\right)$ est dans la périodicité annuelle plus du double de celle observée pour la périodicité hebdomadaire (0,28 contre 0,12 et 0,13). Il en est de même de l'écartement moyennal $\left(\frac{d_m}{v_m}\right.$ étant de 8 p. 100 pour la périodicité annuelle et de 3 p. 100 pour la périodicité hebdomadaire et $\frac{\sigma}{v_m}$ étant respectivement de 9 et de 4 p. 100$\left.\right)$. L'écartement des termes successifs est aussi pour la périodicité annuelle plus fort que pour la périodicité hebdomadaire $\left(\frac{e_m}{v_m}\right.$ est respectivement de 6,3 et de 4,0 p. 100 et $\frac{\varsigma}{v_m}$ est de 7,5 et de 5,6 p. 100$\left.\right)$. D'un autre côté, la périodicité annuelle est beaucoup moins concentrée, son coefficient de concentration c_{12} (= 0,11) étant de deux à trois fois plus faible que celui de la périodicité hebdomadaire (0,26 à l'époque de la hausse et 0,33 à l'époque de la baisse de la consommation du gaz ; la différence sous ce rapport est fortement atténuée par le coefficient k_n). La périodicité annuelle est aussi moins asymétrique (ayant d'une façon générale des coefficients de symétrie bien plus élevés). C'est que la périodicité hebdomadaire se trouve déterminée surtout par le caractère en quelque sorte exceptionnel de deux moments de la série (la journée de dimanche et, par anticipation, celle de samedi) tandis que la périodicité annuelle est le résultat de l'action incessante, bien qu'indirecte, d'un facteur continu — la rotation de la terre autour du soleil.

ANNEXE I.

Tables des coefficients pour séries à 12, à 24 et à 7 termes.

TABLE I, *A*.

Coefficients de concentration (c) *en fonction de* $\frac{\sigma}{d_m}$ *et coefficients de précipitation* (p) *des séries périodiques monocéphales en fonction de* $\frac{\varsigma}{e_m}$.

$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7	$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7	$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7
1,00	0,00	0,00	0,00	1,20	0,14	0,08	0,23	1,40	0,28	0,16	0,46
1,01	0,01	0,00	0,01	1,21	0,14	0,09	0,24	1,41	0,28	0,17	0,47
1,02	0,01	0,01	0,02	1,22	0,15	0,09	0,25	1,42	0,29	0,17	0,48
1,03	0,02	0,01	0,03	1,23	0,16	0,09	0,26	1,43	0,30	0,17	0,49
1,04	0,03	0,02	0,05	1,24	0,17	0,10	0,28	1,44	0,30	0,18	0,51
1,05	0,03	0,02	0,06	1,25	0,17	0,10	0,29	1,45	0,31	0,18	0,52
1,06	0,04	0,02	0,07	1,26	0,18	0,11	0,30	1,46	0,32	0,19	0,53
1,07	0,05	0,03	0,8	1,27	0,19	0,11	0,31	1,47	0,32	0,19	0,54
1,08	0,06	0,03	0,9	1,28	0,19	0,11	0,32	1,48	0,33	0,19	0,55
1,09	0,06	0,04	0,10	1,29	0,20	0,12	0,33	1,49	0,34	0,20	0,56
1,10	0,07	0,04	0,11	1,30	0,21	0,12	0,34	1,50	0,34	0,20	0,57
1,11	0,08	0,04	0,13	1,31	0,21	0,13	0,36	1,51	0,35	0,21	0,59
1,12	0,08	0,05	0,14	1,32	0,22	0,13	0,37	1,52	0,36	0,21	0,60
1,13	0,09	0,05	0,15	1,33	0,23	0,13	0,38	1,53	0,37	0,22	0,61
1,14	0,10	0,06	0,16	1,34	0,23	0,14	0,39	1,54	0,37	0,22	0,62
1,15	0,10	0,06	0,17	1,35	0,24	0,14	0,40	1,55	0,38	0,22	0,63
1,16	0,11	0,06	0,18	1,36	0,25	0,15	0,41	1,56	0,39	0,23	0,64
1,17	0,12	0,07	0,20	1,37	0,26	0,15	0,43	1,57	0,39	0,23	0,66
1,18	0,12	0,07	0,21	1,38	0,26	0,15	0,44	1,58	0,40	0,24	0,67
1,19	0,13	0,08	0,22	1,39	0,27	0,16	0,45	1,59	0,41	0,24	0,68
1,20	0,14	0,08	0,23	1,40	0,28	0 16	0,46	1,60	0,41	0,24	0,69

TABLE I, *A*, (*Suite*).

$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7	$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7	$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}	$\frac{\sigma}{d_m}$ ou $\frac{\varsigma}{e_m}$	c_{12} ou p_{12}	c_{24} ou p_{24}
1,60	0,41	0,24	0,69	1,80	0,55	0,32	0,92	2,00	0,69	0,41	2,20	0,83	0,49
1,61	0,42	0,25	0,70	1,81	0,56	0,33	0,93	2,01	0,70	0,41	2,21	0,83	0,49
1,62	0,43	0,25	0,71	1,82	0,57	0,33	0,94	2,02	0,70	0,41	2,22	0,84	0,50
1,63	0,43	0,26	0,72	1,83	0,57	0,34	0,95	2,03	0,71	0,42	2,23	0,85	0,50
1,64	0,44	0,26	0,74	1,84	0,58	0,34	0,97	2,04	0,72	0,42	2,24	0,86	0,50
1,65	0,45	0,26	0,75	1,85	0,59	0,34	0,98	2,05	0,72	0,43	2,25	0,86	0,51
1,66	0,46	0,27	0,76	1,86	0,59	0,35	0,99	2,06	0,73	0,43	2,26	0,87	0,51
1,67	0,46	0,27	0,77	1,87	0,60	0,35	1,00	2,07	0,74	0,43	2,27	0,88	0,52
1,68	0,47	0,28	0,78	1,88	0,61	0,36	—	2,08	0,75	0,44	2,28	0,88	0,52
1,69	0,48	0,28	0,79	1,89	0,61	0,36	—	2,09	0,75	0,44	2,29	0,89	0,52
1,70	0,48	0,28	0,80	1,90	0,62	0,37	—	2,10	0,76	0,45	2,30	0,90	0,53
1,71	0,49	0,29	0,82	1,91	0,63	0,37	—	2,11	0,77	0,45	2,31	0,90	0,53
1,72	0,50	0,29	0,83	1,92	0,63	0,37	—	2,12	0,77	0,45	2,32	0,91	0,54
1,73	0,50	0,30	0,84	1,93	0,64	0,38	—	2,13	0,78	0,46	2,33	0,92	0,54
1,74	0,51	0,30	0,85	1,94	0,65	0,38	—	2,14	0,79	0,46	2,34	0,92	0,54
1,75	0,52	0,30	0,86	1,95	0,66	0,39	—	2,15	0,79	0,47	2,35	0,93	0,55
1,76	0,52	0,31	0,87	1,96	0,66	0,39	—	2,16	0,80	0,47	2,36	0,94	0,55
1,77	0,53	0,31	0,89	1,97	0,67	0,39	—	2,17	0,81	0,47	2,37	0,95	0,56
1,78	0,54	0,32	0,90	1,98	0,68	0,40	—	2,18	0,81	0,48	2,38	0,95	0,56
1,79	0,55	0,32	0,91	1,99	0,68	0,40	—	2,19	0,82	0,48	2,39	0,96	0,56
1,80	0,55	0,32	0,92	2,00	0,69	0,41	—	2,20	0,83	0,49	2,40	0,97	0,57

TABLE I, *A*, (*suite et fin*).

$\frac{\sigma}{dm}$ ou $\frac{\varsigma}{em}$	c_{12} ou p_{12}	c_{24} ou p_{24}	$\frac{\sigma}{dm}$ ou $\frac{\varsigma}{em}$	c_{24} ou p_{24}	$\frac{\sigma}{dm}$ ou $\frac{\varsigma}{em}$	c_{24} ou p_{24}	$\frac{\sigma}{dm}$ ou $\frac{\varsigma}{em}$	c_{24} ou p_{24}	$\frac{\sigma}{dm}$ ou $\frac{\varsigma}{em}$	c_{24} ou p_{24}	$\frac{\sigma}{dm}$ ou $\frac{\varsigma}{em}$	c_{24} ou p_{24}
2,40	0,97	0,57	2,60	0,65	2,80	0,73	3,00	0,81	3,20	0,89	3,40	0,97
2,41	0,97	0,57	2,61	0,65	2,81	0,73	3,01	0,82	3,21	0,90	3,41	0,98
2,42	0,98	0,58	2,62	0,66	2,82	0,74	3,02	0,82	3,22	0,90	3,42	0,98
2,43	0,99	0,58	2,63	0,66	2,83	0,74	3,03	0,82	3,23	0,90	3,43	0,99
2,44	1,00	0,58	2,64	0,67	2,84	0,75	3,04	0,83	3,24	0,91	3,44	0,99
2,45	—	0,59	2,65	0,67	2,85	0,75	3,05	0,83	3,25	0,91	3,45	0,99
2,46	—	0,59	2,66	0,67	2,86	0,75	3,06	0,84	3,26	0,92	3,46	1,00
2,47	—	0,60	2,67	0,68	2,87	0,76	3,07	0,84	3,27	0,92		
2,48	—	0,60	2,68	0,68	2,88	0,76	3,08	0,84	3,28	0,93		
2,49	—	0,60	2,69	0,69	2,89	0,77	3,09	0,85	3,29	0,93		
2,50	—	0,61	2,70	0,69	2,90	0,77	3,10	0,85	3,30	0,93		
2,51	—	0,61	2,71	0,69	2,91	0,78	3,11	0,86	3,31	0,94		
2,52	—	0,62	2,72	0,70	2,92	0,78	3,12	0,86	3,32	0,94		
2,53	—	0,62	2,73	0,70	2,93	0,78	3,13	0,86	3,33	0,95		
2,54	—	0,62	2,74	0,71	2,94	0,79	3,14	0,87	3,34	0,95		
2,55	—	0,63	2,75	0,71	2,95	0,79	3,15	0,87	3,35	0,95		
2,56	—	0,63	2,76	0,71	2,96	0,80	3,16	0,88	3,36	0,96		
2,57	—	0,64	2,77	0,72	2,97	0,80	3,17	0,88	3,37	0,96		
2,58	—	0,64	2,78	0,72	2,98	0,80	3,18	0,88	3,38	0,97		
2,59	—	0,65	2,79	0,73	2,99	0,81	3,19	0,89	3,39	0,97		
2,60	—	0,65	2,80	0,73	3,00	0,81	3,20	0,89	3,40	0,97		

TABLE I, *B*.

Coefficients de concentration (c) *en fonction de* $\frac{\Sigma\, d^2}{\Sigma^2\, d}$ *et coefficients de précipitation* (p) *des séries périodiques monocéphales en fonction de* $\frac{\Sigma\, e^2}{4\,\alpha^2}$.

$\frac{\Sigma d^2}{\Sigma^2 d}$ ou $\frac{\Sigma e^2}{4\,\alpha^2}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7	$\frac{\Sigma d^2}{\Sigma^2 d}$ ou $\frac{\Sigma e^2}{4\,\alpha^2}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7	$\frac{\Sigma d^2}{\Sigma^2 d}$ ou $\frac{\Sigma e^2}{4\,\alpha^2}$	c_{12} ou p_{12}	c_{24} ou p_{24}	c_7 ou p_7
				0,20	0,38	0,48	0,21	0,40	0,82	0,85	0,77
				0,21	0,41	0,51	0,24	0,41	0,84	0,87	0 80
				0,22	0,43	0,53	0,28	0,42	0,86	0,88	0,82
				0,23	0,46	0,55	0,31	0,43	0,88	0,90	0,84
0,04	—	0	—	0,24	0,48	0,57	0,34	0,44	0,90	0,91	0,87
0,05	—	0,04	—	0,25	0,51	0,59	0,37	0,45	0,91	0,93	0,89
0,06	—	0,08	—	0,26	0,53	0,61	0,40	0,46	0,93	0,94	0,91
0,07	—	0,12	—	0,27	0,55	0,63	0,43	0,47	0,95	0,96	0,93
0,08	0	0,16	—	0,28	0,57	0,65	0,46	0,48	0,97	0,97	0,96
0,09	0,03	0,19	—	0,29	0,60	0,66	0,49	0,49	0,98	0.99	0,98
0,10	0,07	0,22	—	0,30	0,62	0,68	0,52	0,50	1,00	1,00	1,00
0,11	0,10	0,25	—	0,31	0,64	0,70	0,54				
0,12	0,14	0,28	—	0,32	0,66	0,72	0,57				
0,13	0,17	0,31	—	0,33	0,68	0,74	0,60				
0,14	0,20	0,34	0	0,34	0,70	0,75	0,62				
0,15	0,24	0,36	0,03	0,35	0,72	0,77	0,65				
0,16	0,27	0,39	0,07	0,36	0,74	0,79	0,67				
0,17	0,30	0,41	0,10	0,37	0,76	0,80	0,70				
0,18	0,32	0,44	0,14	0,38	0,78	0,82	0,72				
0,19	0,35	0,46	0,17	0,39	0,80	0,84	0,75				
0,20	0,38	0,48	0,21	0,40	0,82	0,85	0,77				

TABLE II, *A*.

Coefficients de nivellement (v) *en fonction de* $\frac{d_m}{\sigma}$ *et coefficients de gradualité* (g) *des series périodiques monocéphales en fonction de* $\frac{e_m}{\varsigma}$.

$\frac{d_m}{\sigma}$ ou $\frac{e_m}{\varsigma}$	v_{24} ou g_{24}	$\frac{d_m}{\sigma}$ ou $\frac{e_m}{\varsigma}$	v_{12} ou g_{12}	v_{24} ou g_{24}	v_7 ou g_7	$\frac{d_m}{\sigma}$ ou $\frac{e_m}{\varsigma}$	v_{12} ou g_{12}	v_{24} ou g_{24}	v_7 ou g_7	$\frac{d_m}{\sigma}$ ou $\frac{e_m}{\varsigma}$	v_{12} ou g_{12}	v_{24} ou g_{24}	v_7 ou g_7
		0,40	—	0,16	—	0,60	0,32	0,44	0,14	0,80	0,66	0,72	0,57
		0,41	0	0,17	—	0,61	0,34	0,45	0,16	0,81	0,68	0,73	0,59
		0,42	0,02	0,18	—	0,62	0,36	0,47	0,18	0,82	0,70	0,75	0,61
		0,43	0,04	0,20	—	0,63	0,38	0,48	0,21	0,83	0,71	0,76	0,63
		0,44	0,05	0,21	—	0,64	0,39	0,49	0,23	0,84	0,73	0,78	0,66
		0,45	0,07	0,23	—	0,65	0,41	0,51	0,25	0,85	0,75	0,79	0,68
		0,46	0,09	0,24	—	0,66	0,43	0,52	0,27	0,86	0,76	0,80	0,70
		0,47	0,10	0,25	—	0,67	0,44	0,54	0,29	0,87	0,78	0,82	0,72
		0,48	0,12	0,27	—	0,68	0,46	0,55	0,31	0,88	0,80	0,83	0,74
0,29	0	0,49	0,14	0,28	—	0,69	0,48	0,56	0,33	0,89	0,81	0,85	0,76
0,30	0,02	0,50	0,16	0,30	—	0,70	0,49	0,58	0,36	0,90	0,83	0,86	0,79
0,31	0,03	0,51	0,17	0,31	—	0,71	0,51	0,59	0,38	0,91	0,85	0,87	0,81
0,32	0,04	0,52	0,19	0,33	—	0,72	0,53	0,61	0,40	0,92	0,87	0,89	0,83
0,33	0,06	0,53	0,21	0,34	—	0,73	0,54	0,62	0,42	0,93	0,88	0,90	0,85
0,34	0,07	0,54	0,22	0,35	0,01	0,74	0,56	0,63	0,44	0,94	0,90	0,92	0,87
0,35	0,09	0,55	0,24	0,37	0,03	0,75	0,58	0,65	0,46	0,95	0,92	0,93	0,89
0,36	0,10	0,56	0,26	0,38	0,05	0,76	0,59	0,66	0,48	0,96	0,93	0,94	0,91
0,37	0,11	0,57	0,27	0,40	0,08	0,77	0,61	0,68	0,51	0,97	0,95	0,96	0,94
0,38	0,13	0,58	0,29	0,41	0,10	0,78	0,63	0,69	0,53	0,98	0,97	0,97	0,96
0,39	0,14	0,59	0,31	0,42	0,12	0,79	0,65	0,70	0,55	0,99	0,98	0,99	0,98
0,40	0,16	0,60	0,32	0,44	0,14	0,80	0,66	0,72	0,57	1,00	1,00	1,00	1,00

TABLE II, *B*.

Coefficients de nivellement (ν) *en fonction de* $\frac{\Sigma^2 d}{\Sigma d^2}$ *et coefficients de gradualité* (g) *des series périodiques monocéphales eu fonction de* $\frac{4\,\alpha^2}{\Sigma e^2}$.

$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\,\alpha^2}{\Sigma e^2}$	ν_{12} ou g_{12}	ν_{24} ou g_{24}	ν_7 ou g_7	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\,\alpha^2}{\Sigma e^2}$	ν_{12} ou g_{12}	ν_{24} ou g_{24}	ν_7 ou g_7	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\,\alpha^2}{\Sigma e^2}$	ν_{12} ou g_{12}	ν_{24} ou g_{24}	ν_7 ou g_7	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\,\alpha^2}{\Sigma e^2}$	ν_{12} ou g_{12}	ν_{24} ou g_{24}
2,0	0	0	0	4,0	0,29	0,17	0,47	6,0	0,50	0,30	0,84	8,0	0,69	0,40
2,1	0,02	0,01	0,03	4,1	0,30	0,17	0,49	6,1	0,51	0,30	0,86	8,1	0,70	0,41
2,2	0,03	0,02	0,06	4,2	0,31	0,18	0,51	6,2	0,52	0,31	0,87	8,2	0,71	0,41
2,3	0,05	0,03	0,08	4,3	0,32	0,19	0,53	6,3	0,53	0,31	0,89	8,3	0,72	0,42
2,4	0,07	0,04	0,11	4,4	0,33	0,19	0,55	6,4	0,54	0,32	0,90	8,4	0,73	0,42
2,5	0,08	0,05	0,13	4,5	0,35	0,20	0,57	6,5	0,55	0,32	0,92	8,5	0,73	0,43
2,6	0,10	0,06	0,16	4,6	0,36	0,21	0,59	6,6	0,56	0,33	0,94	8,6	0,74	0,43
2,7	0,11	0,06	0,19	4,7	0,37	0,22	0,61	6,7	0,57	0,34	0,95	8,7	0,75	0,44
2,8	0,13	0,07	0,21	4,8	0,38	0,22	0,63	6,8	0,58	0,34	0,97	8,8	0,76	0,44
2,9	0,14	0,08	0,23	4,9	0,39	0,23	0,65	6,9	0,59	0,35	0,98	8,9	0,77	0,45
3,0	0,16	0,09	0,26	5,0	0,40	0,24	0,67	7,0	0,60	0,35	1,00	9,0	0,77	0,45
3,1	0,17	0,10	0,28	5,1	0,41	0,24	0,68	7,1	0,61	0,36	—	9,1	0,78	0,46
3,2	0,18	0,11	0,30	5,2	0,42	0,25	0,70	7,2	0,62	0,36	—	9,2	0,79	0,46
3,3	0,20	0,11	0,33	5,3	0,43	0,25	0,72	7,3	0,63	0,37	—	9,3	0,80	0,47
3,4	0,21	0,12	0,35	5,4	0,44	0,26	0,74	7,4	0,64	0,37	—	9,4	0,81	0,47
3,5	0,22	0,13	0,37	5,5	0,45	0,27	0,76	7,5	0,65	0,38	—	9,5	0,81	0,48
3,6	0,24	0,14	0,39	5,6	0,47	0,27	0,77	7,6	0,66	0,38	—	9,6	0,82	0,48
3,7	0,25	0,15	0,41	5,7	0,48	0,28	0,79	7,7	0,66	0,39	—	9,7	0,83	0,49
3,8	0,26	0,15	0,43	5,8	0,49	0,29	0,81	7,8	0,67	0,39	—	9,8	0,84	0,49
3,9	0,27	0,16	0,45	5,9	0,50	0,29	0,82	7,9	0,68	0,40	—	9,9	0,85	0,50
4,0	0,29	0,17	0,47	6,0	0,50	0,30	0,84	8,0	0,69	0,40	—	10,0	0,85	0,50

TABLE II, *B.* (*suite et fin*).

$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{12} ou g_{12}	ν_{24} ou g_{24}	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{24} ou g_{24}	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{24} ou g_{24}	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{24} ou g_{24}	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{24} ou g_{24}	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{24} ou g_{24}	$\frac{\Sigma^2 d}{\Sigma d^2}$ ou $\frac{4\alpha^2}{\Sigma e^2}$	ν_{24} ou g_{24}
10,0	0,85	0,50	12,0	0,59	14,0	0,67	16,0	0,74	18,0	0,81	20,0	0,88	22,0	0,94
10,1	0,86	0,51	12,1	0,59	14,1	0,67	16,1	0,75	18,1	0,81	20,1	0,88	22,1	0,94
10,2	0,87	0,51	12,2	0,60	14,2	0,68	16,2	0,75	18,2	0,82	20,2	0,88	22,2	0,95
10,3	0,88	0,51	12,3	0,60	14,3	0,68	16,3	0,75	18,3	0,82	20,3	0,89	22,3	0,95
10,4	0,88	0,52	12,4	0,60	14,4	0,68	16,4	0,76	18,4	0,82	20,4	0,89	22,4	0,95
10,5	0,89	0,52	12,5	0,61	14,5	0,69	16,5	0,76	18,5	0,83	20,5	0,89	22,5	0,96
10,6	0,90	0,53	12,6	0,61	14,6	0,69	16,6	0,76	18,6	0,83	20,6	0,90	22,6	0,96
10,7	0,91	0,53	12,7	0,62	14,7	0,69	16,7	0,77	18,7	0,83	20,7	0,90	22,7	0,96
10,8	0,91	0,54	12,8	0,62	14,8	0,70	16,8	0,77	18,8	0,84	20,8	0,90	22,8	0,96
10,9	0,92	0,54	12,9	0,62	14,9	0,70	16,9	0,77	18,9	0,84	20,9	0,91	22,9	0,97
11,0	0,93	0,55	13,0	0,63	15,0	0,71	17,0	0,78	19,0	0,84	21,0	0,91	23,0	0,97
11,1	0,94	0,55	13,1	0,63	15,1	0,71	17,1	0,78	19,1	0,85	21,1	0,91	23,1	0,97
11,2	0,94	0,55	13,2	0,64	15,2	0,71	17,2	0,78	19,2	0,85	21,2	0,92	23,2	0,98
11,3	0,95	0,56	13,3	0,64	15,3	0,72	17,3	0,79	19,3	0,85	21,3	0,92	23,3	0,98
11,4	0,96	0,56	13,4	0,64	15,4	0,72	17,4	0,79	19,4	0,86	21,4	0,92	23,4	0,98
11,5	0,96	0,57	13,5	0,65	15,5	0,72	17,5	0,79	19,5	0,86	21,5	0,92	23,5	0,98
11,6	0,97	0,57	13,6	0,65	15,6	0,73	17,6	0,80	19,6	0,86	21,6	0,93	23,6	0,99
11,7	0,98	0,58	13,7	0,66	15,7	0,73	17,7	0,80	19,7	0,87	21,7	0,93	23,7	0,99
11,8	0,99	0,58	13,8	0,66	15,8	0,73	17,8	0,80	19,8	0,87	21,8	0,93	23,8	0,99
11,9	0,99	0,58	13,9	0,66	15,9	0,74	17,9	0,81	19,9	0,87	21,9	0,94	23,9	0,997
12,0	1,00	0,59	14,0	0,67	16,0	0,74	18,0	0,81	20,0	0,88	22,0	0,94	24,0	1,000

ANNEXE II.

Tableaux statistiques.

TABLEAU I.

Données bimensuelles sur les variations de la température de l'air atmosphérique et du sol à 1^m et à 3^m de profondeur à Berlin (moyennes des 5 ans 1910-1914).

D'après les chiffres du *Statistisches Jahrbuch der Stadt Berlin*, vol. 33, Berlin 1916.
(Degrés centigrades).

Mois	Air atmosphérique		Sol profondeur 1m		Sol profondeur 3m	
	le 1er	le 15	le 1er	le 15	le 1er	le 15
Janvier	1,6	1,7	5,6	4,4	9,5	9,0
Février	3,6	4,2	3,1	3,3	8,3	7,8
Mars	6,8	9,0	4,3	5,1	7,5	7,5
Avril	12,5	11,1	6,2	6,6	7,6	8,0
Mai	14,3	18,5	9,2	10,6	8,4	9,1
Juin	21,1	18,7	12,4	14,2	10,0	10,9
Juillet	20,7	25,5	14,7	15,7	11,8	12,3
Août	24,5	18,2	16,6	16,5	13,1	13,6
Septembre	20,2	17,8	16,0	14,7	13,9	13,9
Octobre	15,4	9,4	13,1	11,4	13,7	13,3
Novembre	7,1	7,2	10,2	8,6	12,5	12,0
Décembre	5,0	6,4	6,6	6,4	11,1	10,4
Moyenne de l'année	12,5		9,8		10,63	

Eléments du calcul des divers indices et coefficients.

(D'après les données relatives à la température du sol à 1m de profondeur observée le 15 de chaque mois (v).

Mois	v	d	e	δ	d^2	e^2
I	4,4	— 5,4	— 2,0	+ 1,34	29,16	4,00
II	3,3	— 6,5	— 1,1	+ 2,44	42,25	1,21
III	5,1	— 4,7	+ 1,8	+ 0,64	22,09	3,24
IV	6,6	— 3,2	+ 1,5	— 0,86	10,24	2,25
V	10,6	+ 0,8	+ 4,0	— 3,26	0,64	16,00
VI	14,2	+ 4,4	+ 3,6	+ 0,34	19,36	12,96
VII	15,7	+ 5,9	+ 1,5	+ 1,84	34,81	2,25
VIII	16,5	+ 6,7	+ 0,8	+ 2,64	44,89	0,64
IX	14,7	+ 4,9	— 1,8	+ 0,84	24,01	3,24
X	11,4	+ 1,6	— 3,3	— 2,46	2,56	10,89
XI	8,6	— 1,2	— 2,8	— 2,86	1,44	7,84
XII	6,4	— 3,4	— 2,2	— 0,66	11,56	4,84
Σ	117.5	[48,7]	[26,4]	[20,18]	243,01	69,36
Moyenne	9,8	4,06	2,2	1,68	20,25	5,78

$$\alpha = 16^\circ,5 - 3^\circ,3 = 13^\circ,2 \ ; \ \frac{\alpha}{v_m} = \frac{13^\circ,2}{9^\circ,8} = 1,35$$

$$\sigma = \sqrt{20^\circ,25} = 4^\circ,5 \ ; \ \varsigma = \sqrt{5^\circ,78} = 2^\circ,4$$

$$\frac{d_m}{\sigma} = \frac{4^\circ,06}{4^\circ,50} = 0,90 \ ; \ \frac{e_m}{\varsigma} = \frac{2^\circ,2}{2^\circ,4} = 0,92 \ ; \ K = \frac{20^\circ,18}{48^\circ,7} = 0,415$$

$$\frac{\sigma}{d_m} = \frac{4^\circ,50}{4^\circ,06} = 1,11 \ ; \ \frac{\varsigma}{e_m} = \frac{2^\circ,4}{2^\circ,2} = 1,09$$

$$c_{12} = 0,69\,(1,11 - 1) = 0,08 \ ; \ p_{12} = 0,69\,(1,09 - 1) = 0,06 \ ;$$
$$k_{12} = 0,6\,K = 0,6 \,.\, 0,415 = 0,25$$

et ainsi de suite.

Pour le calcul des coefficients de symétrie, nous procédons de la façon suivante :

Symétrie des écarts moyennaux :

Centre de symétrie maxima : 1 novembre.

Écarts moyennaux des termes équidistants (d' et d'') :

d'	d''	$d'd''$	$(d')^2$	$(d'')^2$
+ 1,6	— 1,2	— 1,92	2,56	1,44
+ 4,9	— 3,4	— 16,66	24,01	11,56
+ 6,7	— 5,4	— 36,18	44,89	29,16
+ 5,9	— 6,5	— 38,35	34,81	42,25
+ 4,4	— 4,7	— 20,68	19,36	22,09
+ 0,8	— 3,2	— 2,56	0,64	10,24

$$\Sigma d'd'' = -116{,}35 \quad ; \quad \Sigma d^2 = 243{,}01$$

$$s_d = \frac{-116{,}35 \times 2}{243{,}01} = -\frac{232{,}70}{243{,}01} = -0{,}958$$

Symétrie des écarts successifs :

Centre de symétrie maxima : 15 août (égal à 16°,5).

e'	e''	$e'e''$	$(e')^2$	$(e'')^2$
— 0,8	— 1,8	+ 1,44	0,64	3,24
— 1,5	— 3,3	+ 4,95	2,25	10,89
— 3,6	— 2,8	+ 10.08	12,96	7,84
— 4,0	— 2,2	+ 8,80	16,00	4,84
— 1,5	— 2,0	+ 3,00	2,25	4,00
— 1,8	— 1,1	+ 1,98	3,24	1.21

$$\Sigma e'e'' = +30{,}25 \quad ; \quad \Sigma e^2 = 69{,}36$$

$$s_e = \frac{30{,}25 \times 2}{69{,}36} = \frac{60{,}50}{69{,}36} = 0{,}872$$

et ainsi de suite.

TABLEAU II.

Décès d'enfants de moins d'un an, par mois, en Suisse (1926-30), *en France* (1926-29) *et dans l'U. R. S. S.* (1926-27) : *moyennes par jour.*

(D'après *Aperçu de la démographie des divers pays du monde* publié par l'Office Permanent de l'Institut International de Statistique, 1931. La Haye, 1932).

Mois	Suisse	France	U. R. S. S.
Janvier	12,2	184,6	2 053
Février	13,1	197,0	2 309
Mars	12,6	199,5	2 393
Avril	11,9	186,3	2 474
Mai	10,5	179,3	2 041
Juin	9,8	159,8	2 155
Juillet	9,2	192,7	3 540
Août	8,9	227,9	4 063
Septembre	8,8	231,5	2 810
Octobre	8,7	180,1	1 991
Novembre	8,9	139,7	1 676
Décembre	9,9	166,6	1 801
Moyenne de l'année	10,36	187,1	2 439

TABLEAU III.

Chômeurs au Royaume-Uni (p. 100 assurés ; 1913-27) *et au Danemark* (p. 100 syndiqués ; 1910-26) *dans l'industrie du bâtiment et spécialement parmi les peintres :* nombres typiques obtenus par le procédé de « médianes élargies ».

(Données empruntées à notre étude *Les fluctuations saisonnières du chômage dans l'industrie du bâtiment de certains pays européens*, Genève, 1929, Bureau International du Travail).

Mois	Royaume-Uni		Danemark	
	Bâtiment	Peintres	Bâtiment	Peintres
Janvier	9,5	20,5	40,2	55,3
Février	9,3	13,9	35,8	46,6
Mars	8,3	7,0	22,5	21,7
Avril	7,1	3,5	8,4	1,0
Mai	6,5	2,8	5,3	1,5
Juin	5,9	4,2	5,3	3,5
Juillet	6,1	5,9	5,2	3,2
Août	6,3	6,3	7,0	5,5
Septembre	6,8	9,4	6,6	6,2
Octobre	7,1	13,6	9,8	14,4
Novembre	8,1	17,9	19,5	37,6
Décembre	8,9	20,0	34,7	54,1
Moyenne de l'année	7,5	10,4	16,7	20,9

TABLEAU IV.

Consommation journalière moyenne de gaz à Genève (en m³) en 1927.

a) *selon les mois :*

Mois	m³	Mois	m³
Janvier	40.885	Juillet	50.558
Février	42.130	Août	49.230
Mars	43.860	Septembre	54.087
Avril	46.366	Octobre	48.025
Mai	53.085	Novembre	45.664
Juin	53.965	Décembre	45.176
En moyenne de l'année			47.774

b) *selon les jours de la semaine :*

Jours	en janvier-février	en mai-juin
Lundi	40.921	52.010
Mardi	42.704	54.014
Mercredi	42.689	55.429
Jeudi	41.971	54.781
Vendredi	42.366	54.596
Samedi	43.108	55.668
Dimanche	38.058	48.916
En moyenne	41.688	53.631

(D'après les données mises aimablement à notre disposition par la Direction de l'usine à gaz de Genève).

P. R. RIDER

The third and fourth moments of the generalized Lexis theory

1. *Introduction.* — Consider the n independent quantities $y_1, \ldots, y_n$. Let $y = (1/n)\,\Sigma_i\, y_i$. Then the mth *moment* of the y's, about their mean value y, is $(1/n)\,\Sigma_i\,(y_i - y)^m$. The first moment is zero. The second moment is called the *variance* of the y's, and its square root is called their *dispersion* or *standard deviation*. The third moment divided by the cube of the standard deviation is termed the *skewness* and the fourth moment divided by the fourth power of the standard deviation is termed the *kurtosis*.

If each y_i is a variable quantity, let us suppose that

$$E\,(y_i) = a_i\,,\ E\,(y_i^2) = A_i\,,\ E\,(y_i^3) = \alpha_i\,,\ E\,(y_i^4) = \mathfrak{a}_i$$

where $E\,(\ \ )$ denotes the expected value of the quantity in the parenthesis. Coolidge * has shown that the expected value of the variance is

$$n^{-1}\left[n^{-1}\,(n-1)\sum_{i=1}^{n}(A_i - a_i^2) + \sum_{i=1}^{n}(a_i - a)^2\right]$$

where $a = (1/n)\,\Sigma_i\, a_i$. This he calls the *fundamental dispersion theorem*. He has also developed a general formula for the variance of a set of statistical observations, which reduces for special cases to the corresponding formula for so-called Bernoulli, Lexis, and Poisson series respectively.

It is the purpose of this paper to derive, by the method employed by Coolidge, formulas for the expected values of the third and fourth moments, and to show how the usual formulas for the third and

---

* J. L. Coolidge, *The dispersion of observations*, « Bulletin of the American Mathematical Society », vol. 27 (1920-21) pp. 439-442. See also Coolidge, *An Introduction to Mathematical Probability*, pp. 66-71.

fourth moments of Bernoulli, Lexis, and Poisson series emerge as special cases.

2. *The fundamental formula for the third moment.* — The third moment of the y's is

$$(1) \quad (1/n) \sum_{i=1}^{n} (y_i - y)^3 = (1/n) \sum_{i=1}^{n} \{[(y_i - a_i) - (y - a)] + (a_i - a)\}^3.$$

From this we get

$$(2) \quad \sum_i (y_i - y)^3 = \sum_i [(y_i - a_i) - (y - a)]^3 + \\ + 3 \sum_i [(y_i - a_i) - (y - a)]^2 (a_i - a) + \\ + 3 \sum_i [(y_i - a_i) - (y - a)] (a_i - a)^2 + \sum_i (a_i - a)^3 .$$

Now let us write the term in the square brackets as follows:

$$(3) \quad (y_i - a_i) - (y - a) = (y_i - a_i) - (1/n) \sum_{j=1}^{n} (y_j - a_j) = \\ = \frac{n-1}{n} (y_i - a_i) - \frac{1}{n} [(y_1 - a_1) + \ldots + (y_{i-1} - a_{i-1}) + \\ + (y_{i+1} - a_{i+1}) + \ldots + (y_n - a_n)] .$$

When this last expression is cubed and summed over all values of i from 1 to n we get

$$(4) \quad \left(\frac{n-1}{n}\right)^3 \sum_i (y_i - a_i)^3 - \frac{3(n-1)^2}{n^3} \sum_{i \pm j} (y_i - a_i)^2 (y_j - a_j) + \\ + \frac{3(n-1)}{n^3} \sum_{i \pm j} (y_i - a_i)(y_j - a_j)^2 - \frac{1}{n^3} \sum_{i \pm j} (y_j - a_j)^3 = \\ = \left[\frac{(n-1)^3}{n^3} - \frac{n-1}{n^3}\right] \sum_i (y_i - a_i)^3 = \frac{(n-1)(n-2)}{n^2} \sum_i (y_i - a_i)^3$$

since the middle pair of terms vanishes on summation. (It is to be noted that the last term in the first member of (4) is to be summed over both i and j, $i \pm j$, and that it is therefore equivalent to $-n^{-3} \Sigma_i (y_i - a_i)^3$ occurring $n - 1$ times). The expected value of the last expression in (4) is $n^{-2} (n-1)(n-2) \Sigma_i (\alpha_i - 3 A_i a_i + 2 a_i^3)$.

The expression $3\,[(y_i - a_i) - (y - a)]^2\,(a_i - a)$ involved in the second term on the right side of (2) may be written, by using (3),

$$(5)\qquad 3\,n^{-2}\,[(n-1)^2\,(y_i - a_i)^2 - 2\,(n-1)\,(y_i - a_i)\,(y_j - a_j) + \\ + (y_j - a_j)^2]\,(a_i - a)\,.$$

Summing this up, we get

$$3\,n^{-2}[(n-1)^2\,(y_1 - a_1)^2\,(a_1 - a) + \ldots + (n-1)^2\,(y_n - a_n)^2\,(a_n - a)] + \\ + (y_1 - a_1)^2\,(a_2 - a) + \ldots + (y_1 - a_1)^2\,(a_n - a) + \\ + \ldots\ldots\ldots\ldots + \\ + [(y_n - a_n)^2\,(a_1 - a) + \ldots + (y_n - a_n)^2\,(a_{n-1} - a)] - \\ - 2\,n^{-2}\,(n-1) \sum_{i \pm j} (y_i - a_i)\,(y_j - a_j)\,(a_i - a) =$$

$$= 3\,n^{-2} \{ (y_1 - a_1)^2\,[(n-1)^2\,(a_1 - a) + (a_2 - a) + \ldots + (a_n - a)] + \\ + \ldots\ldots\ldots\ldots + \\ + (y_n - a_n)^2\,[(n-1)^2\,(a_n - a) + (a_1 - a) + \ldots + (a_{n-1} - a)] \} - \\ - 2\,n^{-2}\,(n-1) \sum_{i \pm j} (y_i - a_i)\,(y_j - a_j)\,(a_i - a) =$$

$$= 3\,n^{-2}[n\,(n-2)\,(y_1 - a_1)^2\,(a_1 - a) + \ldots + n\,(n-2)\,(y_n - a_n)^2\,(a_n - a)] - \\ - 2\,n^{-2}\,(n-1) \sum_{i \pm j} (y_i - a_i)\,(y_j - a_j)\,(a_i - a) =$$

$$= 3\,n^{-1}\,(n-2) \sum_i (y_i - a_i)^2\,(a_i - a) - \\ - 2\,n^{-2}\,(n-1) \sum_{i \pm j} (y_i - a_i)\,(y_j - a_j)\,(a_i - a)\,.$$

The expected value of this is

$$(6)\qquad 3\,n^{-1}\,(n-2) \sum_i (A_i - a_i^2)\,(a_i - a)\,.$$

The expected value of the third term on the right side of (2) is zero, and the expected value of the last term is its ostensible value. Thus, *the expected value of the third moment is*

$$(7)\qquad n^{-3}\,(n-1)\,(n-2) \sum_{i=1}^{n} (\alpha_i - 3\,A_i\,a_i + 2\,a_i^3) + \\ + 3\,n^{-2}\,(n-2) \sum_{i=1}^{n} (A_i - a_i^2)\,(a_i - a) + n^{-1} \sum_{i=1}^{n} (a_i - a)^3\,.$$

3. *The fundamental formula for the fourth moment.* — The fourth moment of the set of y's is

$$(8) \quad (1/n) \sum_{i=1}^{n} (y_i - y)^4 = (1/n) \sum_{i=1}^{n} \{ [(y_i - a_i) - (y - a)] + (a_i - a) \}^4$$

from which we get

$$(9) \quad \sum_i (y_i - y)^4 = \sum_i [(y_i - a_i) - (y - a)]^4 + $$
$$+ 4 \sum_i [(y_i - a_i) - (y - a)]^3 (a_i - a) + $$
$$+ 6 \sum_i [(y_i - a_i) - (y - a)]^2 (a_i - a)^2 + $$
$$+ 4 \sum_i [(y_i - a_i) - (y - a)] (a_i - a)^3 + \sum_i (a_i - a)^4 .$$

By using (3) we find that

$$(10) \quad [(y_i - a_i) - (y - a)]^4 = \left(\frac{n-1}{n}\right)^4 (y_i - a_i)^4$$
$$- \frac{4(n-1)^3}{n^4} (y_i - a_i)^3 [(y_1 - a_1) + \ldots + (y_{i-1} - a_{i-1}) + $$
$$+ (y_{i+1} - a_{i+1}) + \ldots + (y_n - a_n)] + $$
$$+ \frac{6(n-1)^2}{n^4} (y_i - a_i)^2 [(y_1 - a_1) + \ldots + (y_{i-1} - a_{i-1}) + $$
$$+ (y_{i+1} - a_{i+1}) + \ldots + (y_n - a_n)]^2 - $$
$$- \frac{4(n-1)}{n^4} (y_i - a_i) [(y_1 - a_1) + \ldots + (y_{i-1} - a_{i-1}) + $$
$$+ (y_{i+1} - a_{i+1}) + \ldots + (y_n - a_n)]^3 + $$
$$+ \frac{1}{n^4} [(y_1 - a_1) + \ldots + (y_{i-1} - a_{i-1}) + (y_{i+1} - a_{i+1}) + \ldots + (y_n - a_n)]^4 .$$

The sum of this with respect to i is equal to

$$\left(\frac{n-1}{n}\right)^4 \sum_i (y_i - a_i)^4 + \frac{6(n-1)^2}{n^4} \sum_{i \pm j} (y_i - a_i)^2 (y_j - a_j)^2 + $$
$$+ \frac{n-1}{n^4} \sum_i (y_i - a_i)^4 + \frac{3(n-2)}{n^4} \sum_{i \pm j} (y_i - a_i)^2 (y_j - a_j)^2 + $$

+ terms whose expected value is zero.

The expected value of this expression is

$$(11)\quad n^{-4}\Big\{\Big[(n-1)^4+(n-1)\Big]\sum_i(\mathrm{a}_i-4\,\alpha_i\,a_i+6\,A_i\,a_i^2-3\,a_i^3)+$$

$$+\,[6\,(n-1)^2+3\,(n-2)]\sum_{i\pm j}(A_i-a_i^2)\,(A_j-a_j^2)=$$

$$=n^{-3}\,(n-1)\,(n^2-3\,n+3)\sum_i(\mathrm{a}_i-4\,\alpha_i\,a_i+6\,A_i\,a_i^2-3\,a_i^3)+$$

$$+\,3\,n^{-3}\,(2\,n-3)\Big\{\Big[\sum_i(A_i-a_i^2)-\sum_i(A_i-a_i^2)^2\Big]\Big\}\,.$$

The expected value of the term $4\,\Sigma_i\,[(y_i-a_i)-(y-a)]^3\,(a_i-a)$ in (9) is, after some reductions similar to those used in getting from (5) to (6), found to be

$$(12)\qquad 4\,n^{-2}\,(n^2-3\,n+3)\sum_i(\alpha_i-3\,A_i\,a_i^2+2\,a_i^3)\,(a_i-a)\,.$$

We may use (3) to expand the third expression on the right side of (9) as follows:

$$6\sum_i[(y_i-a_i)-(y-a)]^2\,(a_i-a)^2=$$

$$=6\,\{n^{-1}\,(n-1)\,(y_1-a_1)-n^{-1}\,[(y_2-a_2)+\ldots+(y_n-a_n)]\}^2\,(a_1-a)^2$$

$$+\;.$$

$$+\,6\,\{n^{-1}\,(n-1)\,(y_n-a_n)-n^{-1}\,[(y_1-a_1)+\ldots+$$

$$+\,(y_{n-1}-a_{n-1})]\}^2\,(a_n-a)^2=$$

$$=6\,n^{-2}\,\Big\{(n-1)^2\sum_i(y_i-a_i)^2\,(a_i-a)^2+$$

$$+\,(y_2-a_2)^2\,(a_1-a)^2+\ldots+(y_n-a_n)^2\,(a_1-a)^2+$$

$$+\,(y_1-a_1)^2\,(a_2-a)^2+(y_3-a_3)^2\,(a_2-a)^2+\ldots+(y_n-a_n)^2\,(a_2-a)^2$$

$$+\ldots+(y_1-a_1)^2\,(a_n-a)^2+\ldots+(y_{n-1}-a_{n-1})^2\,(a_n-a)^2\Big\}$$

$+$ terms whose expected value is zero.

The expected value of this is

$$6\,n^{-2}\,\Big\{(n-1)^2\sum_i(A_i-a_i^2)\,(a_i-a)^2+$$

$$+\,[(A_2-a_2^2)+\ldots+(A_n-a_n^2)]\,(a_1-a)^2+$$

$$+\,[(A_1-a_1^2)+(A_3-a_3^2)+\ldots+(A_n-a_n^2)]\,(a_2-a)^2+$$

$$+ \dots + [(A_1 - a_1^2) + \dots + (A_{n-1} - a_{n-1}^2)] (a_n - a)^2 \Big\} =$$

$$= 6\, n^{-2} \Big\{ (n-1)^2 \sum_i (A_i - a_i^2)(a_i - a)^2 +$$

$$+ (A_1 - a_1^2) \sum_i (a_i - a)^2 - (A_1 - a_1^2)(a_1 - a)^2 + \dots +$$

$$+ (A_n - a_n^2) \sum_i (a_i - a)^2 - (A_n - a_n^2)(a_n - a)^2 \Big\} =$$

$$(13) \quad = 6\, n^{-2} \Big\{ n(n-2) \sum_i (A_i - a_i^2)(a_i - a) + \sum_i (A_i - a_i^2) \sum_i (a_i - a^2) \Big\}.$$

The next to the last term in (9) has the expected value zero, and the expected value of the last term is its ostensible value. Making use of these two facts and of (11), (12), (13) we find that *the expected value of the fourth moment is*

$$(14) \quad n^{-4}(n-1)(n^2 - 3n + 3) \sum_{i=1}^{n} (a_i - 4\,\alpha_i\, a_i + 6\, A_i\, a_i^2 - 3\, a_i^4) +$$

$$+ 3\, n^{-4}(2n-3) \Big\{ \Big[\sum_{i=1}^{n} (A_i - a_i^2) \Big]^2 - \sum_{i=1}^{n} (A_i - a_i^2)^2 \Big\} +$$

$$+ 4\, n^{-3}(n_2 - 3n + 3) \sum_{i=1}^{n} (\alpha_i - 3\, A_i\, a_i + 2\, a_i^3)(a_i - a) +$$

$$+ 6\, n^{-2}(n-2) \sum_{i=1}^{n} (A_i - a_i^2)(a_i - a)^2 +$$

$$+ 6\, n^{-3} \sum_{i=1}^{n} (A_i - a_i^2) \sum_{i=1}^{n} (a_i - a)^2 + n^{-1} \sum_{i=1}^{n} (a_i - a)^4 .$$

4. *Application to the Lexis theory.* — Let us now multiply (7) and (14) by n and then replace such expressions as $n^{-k}(n^l + c_1 n^{l-1} + \dots + c_k)$ by unity when $k = l$ and by zero when $k > l$.* We are then led to the approximate equations

$$(15) \qquad E \sum_i (y_i - y)^3 = \sum_i [(\alpha_i - 3\, A_i\, a_i + 2\, a_i^3) +$$

$$+ 3\,(A_i - a_i^2)(a_i - a) + (a_i - a)^3] .$$

* H. L. RIETZ has avoided this approximation in considering the variance of observations. (*On the Lexis theory and the analysis of variance* « Bulletin of the American Mathematical Society », vol. 38 (1932), pp. 731-735). It is sufficient, however, for the purposes of the present paper to use this approximation.

$$(16) \qquad E \sum_i (y_i - y)^4 = \sum_i \Big[(a_i - 4\,\alpha_i\, a_i + 6\, A_i\, a_i^2 - 3\, a_i^4) +$$

$$+ 4\,(\alpha_i - 3\, A_i\, a_i + 2\, a_i^3)\,(a_i - a) + 6\,(A_i - a_i^2)\,(a_i - a)^2 + (a_i - a)^4 \Big].$$

To show how these equations may be applied to the Lexis theory, let us suppose that we have N sets, each of s observations:

$$(17) \qquad \begin{aligned} &x_{11}, x_{12}, \ldots, x_{1s}\ ; \ \sum_{j=1}^{s} x_{1j} = x_1 \\ &x_{21}, x_{22}, \ldots, x_{2s}\ ; \ \sum_{j=1}^{s} x_{2j} = x_2 \\ &\ldots\ldots\ldots\ldots\ldots \\ &x_{N1}, x_{N2}, \ldots, x_{Ns}\ ; \ \sum_{j=1}^{s} x_{Nj} = x_N. \end{aligned}$$

Let us suppose further that the probability associated with x_{ij} is p_{ij}, this being the chance that x_{ij} takes the value 1, while in the opposite case it takes the value 0. Then *

$$(18) \qquad \begin{aligned} &E\,(x_{ij}) = p_{ij} \quad , \ E\,(x_{ij} - p_{ij})^2 = p_{ij} - p_{ij}^2 \\ &E\,(x_{ij} - p_{ij})^3 = p_{ij} - 3\, p_{ij}^2 + 2\, p_{ij}^3 \\ &E\,(x_{ij} - p_{ij})^4 = p_{ij} - 4\, p_{ij}^2 + 6\, p_{ij}^3 - 3\, p_{ij}^4 . \end{aligned}$$

Now let

$$(19) \qquad \sum_{j=1}^{s} p_{ij} = s\, p_i \ , \quad \sum_{i=1}^{N} p_i = N\, p \ , \quad \sum_{i=1}^{N} x_i = N\, x .$$

From (19), $E\,(x_i) = s\, p_i$, and since $x_i - s\, p_i = \Sigma_j\,(x_{ij} - p_{ij})$, it follows that

$$(20) \qquad E\,(x_i - s\, p_i)^2 = E \left[\sum_{j=1}^{s} (x_{ij} - p_{ij}) \right]^2 =$$

$$= E \sum_{j=1}^{s} (x_{ij} - p_{ij})^2 = \sum_{j=1}^{s} (p_{ij} - p_{ij}^2)$$

$$(21) \qquad E\,(x_i - s\, p_i)^3 = E \left[\sum_{j=1}^{s} (x_{ij} - p_{ij}) \right]^3 =$$

$$= E \sum_{j=1}^{s} (x_{ij} - p_{ij})^3 = \sum_{j=1}^{s} (p_{ij} - 3\, p_{ij}^2 + 2\, p_{ij}^3) .$$

---

* Cf. T. C. Fry, *Probability and its Engineering Uses*, p. 471.

This last expression corresponds to $\alpha_i - 3 A_i a_i + 2 a_i^3$ in (15).

$$(22) \qquad E(x_i - s p_i)^4 = E\left[\sum_{j=1}^{s} (x_{ij} - p_{ij})\right]^4 =$$

$$= E \sum_{j=1}^{s} (x_{ij} - p_{ij})^4 + 3 E \sum_{j \pm k} (x_{ij} - p_{ij})^2 (x_{ik} - p_{ik})^2 =$$

$$= \sum_{j=1}^{s} (p_{ij} - 4 p_{ij}^2 + 6 p_{ij}^3 - 3 p_{ij}^4) + 3 \sum_{j \pm k} (p_{ij} - p_{ij}^2)(p_{ik} - p_{ik}^2)$$

corresponding to $a_i - 4 \alpha_i a_i + 6 A_i a_i^2 - 3 a_i^4$ in (16).

By (15) and (16), we get, respectively,

$$(23) \qquad E \sum_{i=1}^{N} (x_i - x)^3 = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{s} \left[(p_{ij} - 3 p_{ij}^2 + 2 p_{ij}^3) + \right.\right.$$

$$\left.\left. + 3 s (p_{ij} - p_{ij}^2)(p_i - p) + s^3 (p_i - p)^3 \right] \right\}$$

$$(24) \qquad E \sum_{i=1}^{N} (x_i - x)^4 = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{s} (p_{ij} - 4 p_{ij}^2 + 6 p_{ij}^3 - 3 p_{ij}^4) + \right.$$

$$+ 3 \sum_{j \pm k} (p_{ij} - p_{ij}^2)(p_{ik} - p_{ik}^2) +$$

$$+ \sum_{j=1}^{s} \left[4 s (p_{ij} - 3 p_{ij}^2 + 2 p_{ij}^3)(p_i - p) + \right.$$

$$\left.\left. + 6 s^2 (p_{ij} - p_{ij}^2)(p_i - p)^2 + s^4 (p_i - p)^4 \right] \right\}.$$

To reduce the right sides of these equations to the desired form we need the following formulas, which are not very difficult to derive:

$$(25) \qquad \sum_{j=1}^{s} p_{ij} = s p_i, \quad \sum_{j=1}^{s} p_{ij}^2 = s p_i^2 + \sum_{j=1}^{s} (p_{ij} - p_i)^2$$

$$\sum_{j=1}^{s} p_{ij}^3 = s p_i^3 + 3 p_i \sum_{j=1}^{s} (p_{ij} - p_i)^2 + \sum_{j=1}^{s} (p_{ij} - p_i)^3$$

$$\sum_{j=1}^{s} p_{ij}^4 = s p_i^4 + 6 p_i^2 \sum_{j=1}^{s} (p_{ij} - p_i)^2 + 4 p_i \sum_{j=1}^{s} (p_{ij} - p_i) +$$

$$+ \sum_{j=1}^{s} (p_{ij} - p_i)^4$$

$$\sum_{j \pm k} p_{ij} p_{ik} = s(s-1) p_i^2 - \sum_{j=1}^{s} (p_{ij} - p_i)^2$$

$$\sum_{j \pm k} p_{ij}^2 p_{ik} = s(s-1)\,p_i^3 + (s-3)\,p_i \sum_{j=1}^{s} (p_{ij} - p_i)^2 - \\ - \sum_{j=1}^{s} (p_{ij} - p_i)^3$$

$$\sum_{j \pm k} p_{ij}^2 p_{ik}^2 = s(s-1)\,p_i^4 + 2(s-3)\,p_i^2 \sum_{j=1}^{s} (p_{ij} - p_i)^2 - \\ - 4\,p_i \sum_{j=1}^{s} (p_{ij} - p_i)^3 + \left[\sum_{j=1}^{s} (p_{ij} - p_i)^2\right]^2 - \sum_{j=1}^{s} (p_{ij} - p_i)^4 .$$

Substituting these values in (23) we find

$$(26)\quad E \sum_{i=1}^{N} (x_i - x)^3 = \sum_{i=1}^{N} \Big\{ s\,p_i - 3\,s\,p_i^2 + 2\,s\,p_i^3 + 3\,s^2 p_i^2 (p_i - p) - \\ - 3\,s^2 p_i^2 (p_i - p) + s^3 (p_i - p)^3 - 3\left[(q - p) + (s-2)(p_i - p)\right] \\ \sum_{j=1}^{s} (p_{ij} - p_i)^2 + 2 \sum_{j=1}^{s} (p_{ij} - p_i)^3 \Big\}$$

in which $q = 1 - p$.

A further reduction may be effected by means of the relations

$$(27)\quad \sum_{i=1}^{N} p_i = N\,p\,,\quad \sum_{i=1}^{N} p_i^2 = N\,p^2 + \sum_{i=1}^{N} (p_i - p)^2$$

$$\sum_{i=1}^{N} p_i^3 = N\,p^3 + 3\,p \sum_{i=1}^{N} (p_i - p)^2 + \sum_{i=1}^{N} (p_i - p)^3$$

$$\sum_{i=1}^{N} p_i^4 = N\,p^4 + 6\,p^2 \sum_{i=1}^{N} (p_i - p)^2 + 4\,p \sum_{i=1}^{N} (p_i - p)^3 + \sum_{i=1}^{N} (p_i - p)^4 .$$

Using (27), we can readily reduce (26) to the following form:

$$(28)\quad E \sum_{i=1}^{N} (x_i - x)^3 = N\,s\,p\,q\,(q - p) + 3\,s(s-1)(q-p) \sum_{i=1}^{N} (p_i - p)^2 \\ + s(s-1)(s-2) \sum_{i=1}^{N} (p_i - p)^3 - 3\,(q - p) \sum_{i=1,j=1}^{N,s} (p_{ij} - p_i)^2 \\ - 3\,(s-2) \sum_{i=1}^{N} (p_i - p) \sum_{j=1}^{s} (p_{ij} - p_i)^2 + 2 \sum_{i=1,j=1}^{N,s} (p_{ij} - p_i)^2 .$$

Returning now to (24), into which we substitute (25), we get

$$(29)\quad E\sum_{i=1}^{N}(x_i-x)^4=\sum_{i=1}^{N}\Big\{s\,p_i+[-4\,s+3\,s\,(s-1)]\,p_i^2+$$
$$+[6\,s-6\,s\,(s-1)]\,p_i^3+[-3\,s+3\,s\,(s-1)]\,p_i^4+$$
$$+4\,s^2\,(p_i-p)\,p_i-12\,s^2\,(p_i-p)\,p_i^2+8\,s^2\,(p_i-p)\,p_i^3+$$
$$+6\,s^3\,(p_i-p)^2\,(p_i-p_i^2)+s^4\,(p_i-p)^4+$$
$$+[-4+18\,p_i-18\,p_i^2-3-6\,(s-3)\,p_i+6\,(s-3)\,p_i^2-$$
$$-12\,s\,(p_i-p)+24\,s\,p_i\,(p_i-p)-6\,s^2\,(p_i-p)^2]\sum_{j=1}^{s}(p_{ij}-p_i)^2+$$
$$+[12-24\,p_i+8\,s\,(p_i-p)]\sum_{j=1}^{s}(p_{ij}-p_i)^3-6\sum_{j=1}^{s}(p_{ij}-p_i)^4+$$
$$+3\left[\sum_{j=1}^{s}(p_{ij}-p_i)^2\right]^2\Big\}.$$

Making use of (27), we can reduce the foregoing to

$$(30)\qquad E\sum_{i=1}^{N}(x_i-x)^4=N\,[s\,p\,q\,(1-6\,p\,q)+3\,s^2\,p^2\,q^2]+$$
$$+s\,(s-1)\,[7+6\,(s-6)\,p\,q]\sum_{i=1}^{N}(p_i-p)^2+$$
$$+6\,s\,(s-1)\,(s-2)\,(q-p)\sum_{i=1}^{N}(p_i-p)^3+$$
$$+s\,(s-1)\,(s-2)\,(s-3)\sum_{i=1}^{N}(p_i-p)^4-$$
$$-[7+6\,(s-6)\,p\,q]\sum_{i=1,j=1}^{N,s}(p_{ij}-p_i)^2-$$
$$-6\,(s-2)\sum_{i=1}^{N}[3\,(q-p)\,(p_i-p)+$$
$$+(s-3)\,(p_i-p)^2]\sum_{j=1}^{s}(p_{ij}-p_i)^2+$$
$$+12\,(q-p)\sum_{i=1,j=1}^{N,s}(p_{ij}-p_i)^3+$$
$$+8\,(s-3)\sum_{j=1}^{N}(p_i-p)\sum_{j=1}^{s}(p_{ij}-p_i)^3-$$
$$-6\sum_{i=1,j=1}^{N,s}(p_{ij}-p_i)^4+3\sum_{i=1}^{N}\left[\sum_{j=1}^{s}(p_{ij}-p_i)^2\right]^2.$$

Equations (28) and (30) are our fundamental formulas. Let us see how they reduce for special cases.

Bernoulli Series. — In a Bernoulli series the probability is constant throughout, that is, $p_{ij} = p_i = p$. Equations (28) and (30), after division by N, reduce to

$$N^{-1} E \sum_{i=1}^{N} (x_i - x)^3 = s\, p\, q\, (q - p) \tag{31}$$

$$N^{-1} E \sum_{i=1}^{N} (x_i - x)^4 = s\, p\, q\, (1 - 6\, p\, q) + 3\, s^2 p^2 q^2 \tag{32}$$

which are recognized as the third and fourth moments respectively of a Bernoulli series. *

Lexis series. — Here the probability is constant within a set, but varies from set to set: $p_{ij} = p_i \mp p$. Formulas (28) and (29) become

$$\begin{aligned} N^{-1} E \sum_{i=1}^{N} (x_i - x)^3 &= s\, p\, q\, (q - p) + \\ &+ 3\, s\, (s - 1)\, (q - p)\, N^{-1} \sum_{i=1}^{N} (p_i - p)^2 + \\ &+ s\, (s - 1)\, s - 2)\, N^{-1} \sum_{i=1}^{N} (p_i - p)^3 \end{aligned} \tag{33}$$

$$\begin{aligned} N^{-1} E \sum_{i=1}^{N} (x_i - x)^4 &= s\, p\, q\, (1 - 6\, p\, q) + 3\, s^2 p^2 q^2 + \\ &+ s\, (s - 1)\, [7 + 6\, (s - 6)\, p\, q]\, N^{-1} \sum_{i=1}^{N} (p_i - p)^2 + \\ &+ 6\, s\, (s - 1\, (s - 2)\, (q - p)\, N^{-1} \sum_{i=1}^{N} (p_i - p)^3 + \\ &+ s\, (s - 1)\, (s - 2)\, (s - 3)\, N^{-1} \sum_{i=1}^{N} (p_i - p)^4 . \end{aligned} \tag{34}$$

These are the third and fourth moments of a Lexis series. **

Poisson Series. — In this type of series the probability varies within the set, but all sets are comparable: $p_{ij} \mp p_i = p$. We get from (28) and (30)

---

* See Fry, op. cit., p. 471.

** See Charles Jordan, *Statistique Mathématique*, pp. 113-115.

$$(35) \qquad N^{-1} E \sum_{i=1}^{N} (x_i - x)^3 = s\,p\,q\,(q - p) -$$

$$- 3\,(q - p) \sum_{j=1}^{s} (p_{ij} - p_i)^2 + 2 \sum_{j=1}^{s} (p_{ij} - p_i)^3$$

$$(36) \qquad N^{-1} E \sum_{i=1}^{N} (x_i - x)^4 = s\,p\,q\,(1 - 6\,p\,q) +$$

$$+ 3\,s^2 p^2 q^2 - [7 + 6\,(s - 6\,p\,q]\sum_{j=1}^{s} (p_{ij} - p_i)^2 +$$

$$+ 12\,(q - p) \sum_{j=1}^{s} (p_{ij} - p_i)^3 -$$

$$- 6 \sum_{j=1}^{s} (p_{ij} - p_i)^4 + 3 \left[\sum_{j=1}^{s} (p_{ij} - p_i)^2 \right]^2$$

which are the formulas for the third and fourth moments of a Poisson series. *

5. *Illustration* ** *of the fundamental formulas for the third and fourth moments.* — In order to illustrate the fundamental formulas (7) and (14) for the expected values of the third and fourth moments respectively, let us consider the accompanying table of monthly precipitation in New York City (Table I). For a given year, y_i is the precipitation for the ith month. Thus, for 1900, $y_1 = 4.18$ (January precipitation), $y_2 = 5.16$ (February precipitation), and so on. The variable a_i in the formulas is the mean precipitation for the ith month for the entire period of ten years, A_i the mean square precipitation for the ith month, α_i the mean cube, and a_i the mean fourth power. The values of these quantities are shown in Table II. Substituting them in formula (7), in which n is now of course 12, we find that the expected value of the third moment of the precipitation for a given year is 2.3522976. A similar substitution in (14) yields 31.354688 for the expected value of the fourth moment.

For comparison purposes, let us compute these two moments for a specific year, say 1907. The mean value of the precipitation for this years is $y = 3.69$, and we find

$$\Sigma\,(y_i - y)^3 = (3.26 - 3.69)^3 + \ldots + (3.91 - 3.69)^3 = 54.713310$$
$$\Sigma\,(y_i - y)^4 = (3.26 - 3.69)^4 + \ldots + (3.91 - 3.69)^4 = 412.58839640$$

* Cf. JORDAN, op. cit., pp. 109-110.

** The investigations of sections 5 and 6 were made possible by assistance from a grant made by the *Rockefeller Foundation* to Washington University for Research in Science.

Dividing each of these by 12 we obtain for the third and fourth moments 4.5594425 and 34.382366 respectively.

6. *Example of the application to the generalized Lexis theory.* — As an illustration of the application to the generalized Lexis theory, the following sampling experiment was conducted : Four sets of eight trials each were made to obtain a sample. Five hundred such samples were obtained. The probability of success varied both from trial to trial within a set and from set to set as shown in Table III.

In the notation of section 4, $p_{11} = .02$, $p_{12} = .04$, ... The average probabilities for the four sets are respectively

$$p_1 = .09\,, \quad p_2 = .25\,, \quad p_3 = .28\,, \quad p_4 = .38\,.$$

The general average is $p = \frac{1}{4}(p_1 + p_2 + p_3 + p_4) = .25$, and $q = 1 - p = .75$.

The values of s and N are 8 and 4 respectively, and it is found that

$$\sum_{i=1}^{4} (p_i - p)^2 = 0.0434\,, \quad \sum_{i=1}^{4} (p_i - p)^3 = -0.001872\,,$$

$$\sum_{i=1}^{4} (p_i - p)^4 = 0.00094178\,,$$

$$\sum_{j=1}^{8} (p_{1j} - p_1)^2 = (.02 - .09)^2 + (.04 - .09)^2 + \ldots +$$
$$+ (.16 - .09)^2 = 0.0168\,,$$

$$\sum_{j=1}^{8} (p_{2j} - p_2)^2 = 0.2856\,, \quad \sum_{j=1}^{8} (p_{3j} - p_3)^2 = 0.4336\,,$$

$$\sum_{j=1}^{8} (p_{4j} - p_4)^2 = 0.3968\,,$$

$$\sum_{j=1}^{8} (p_{1j} - p_1)^3 = 0\,, \quad \sum_{j=1}^{8} (p_{2j} - p_2)^3 = 0.032640\,,$$

$$\sum_{j=1}^{8} (p_{3j} - p_3)^3 = 0.042240\,, \quad \sum_{j=1}^{8} (p_{4j} - p_4)^3 = 0.110400\,,$$

$$\sum_{j=1}^{8} (p_{1j} - p_1)^4 = 0.00006216\,, \quad \sum_{j=1}^{8} (p_{2j} - p_2)^4 = 0.02083656\,,$$

$$\sum_{j=1}^{8} (p_{3j} - p_3)^4 = 0.04042816\,, \quad \sum_{j=1}^{8} (p_{4j} - p_4)^4 = 0.06743936\,.$$

Substituting the foregoing values in equations (28) and (30) we find

$$E \sum_{i=1}^{3} (x_i - x)^3 = 3.573696\ , E \sum_{i=1}^{4} (x_i - x)^4 = 33.29714176\ .$$

Division by 4 yields 0.893424 and 8.32428544 for the expected values of the third and fourth moments respectively.

In the actual sampling experiment, which was conducted by means of Tippett's numbers, * the number of successes obtained in the eight trials from set 1 was tabulated, similarly the number of successes in sets 2, 3, and 4 were tabulated. These four numbers constituted a sample. For example, a typical sample consisted of 1 success in the first set of eight trials, 2 successes in the second set, 4 in the third, and 5 in the fourth. This sample therefore was composed of the numbers 1, 2, 4, 5, and the various moments could be computed.

Since we were chiefly interested in the third and fourth moments, these moments were computed for all 500 samples and their mean values were determined to be 0.1005 and 5.9805625 respectively, which may be compared with the theoretical values given above. The frequency distributions of these moments are exhibited in Tables IV and V.

TABLE I.

Monthly Precipitation in Inches, New York City. *

	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909
January . .	4.18	2.07	2.28	3.44	3.38	3.93	2.98	3.26	3.84	3.3
February . .	5.16	0.86	5.78	3.83	2.18	2.79	2.57	1.52	5.36	4.31
March . . .	3.18	5.18	4.32	3.65	3.44	3.65	5.58	3.80	2.15	3.19
April. . . .	2.06	6.82	3.51	2.88	3.94	2.45	5.78	3.89	1.82	5.93
May	4.08	7.01	1.23	0.33	1.61	1.12	4.67	4.08	9.10	1.72
June	3.36	0.94	5.91	7.42	2.70	4.18	1.70	3.29	1.70	3.17
July	4.33	5.41	3.12	3.33	4.31	6.01	3.21	1.18	4.33	1.98
August . . .	2.69	6.88	3.29	5.96	7.13	5.23	3.68	2.48	5.65	7.94
September .	2.36	2.33	3.59	2.60	3.18	7.11	2.54	8.00	1.60	2.66
October . .	4.17	2.20	6.66	1.55	3.21	2.67	4.30	3.82	1.92	0.74
November .	4.26	1.31	1.19	0.90	2.62	1.67	1.28	5.05	0.75	1.58
December. .	1.98	6.05	6.19	2.81	3.87	3.67	3.53	3.91	3.21	5.00

* COOLIDGE, *An introduction to mathematical probability*, p. 69.

* L. H. C. TIPPETT, *Random Sampling Numbers.*

TABLE II.

MONTH	i	a_i	A_i	α_i	a_i
January . .	1	3.269	11.10011	38.8435949	139.196824115
February .	2	3.436	14.42000	67.3924510	334.336897940
March . . .	3	3.814	15.44608	66.0745912	296.844512116
April . . .	4	3.908	17.99144	93.5477762	527.164942004
May	5	3.495	20.54785	134.8288749	1032.142427109
June	6	3.437	15.35704	82.3717133	526.962292323
July	7	3.721	15.79023	72.7629637	355.060233255
August . . .	8	5.093	29.37649	184.4005154	1221.932609866
September .	9	3.597	17.13983	103.2524895	712.431042707
October . .	10	3.124	12.37364	57.7135288	302.469404132
November .	11	2.061	6.19429	23.9864181	104.945005477
December .	12	4.022	17.82340	88.1333292	442.393866488
Total		42.978	193.56040	1013.3082462	5995.890057532

a_i = mean precipitation

A_i = mean square precipitation

α_i = mean cube precipitation

a_i = mean fourth power of precipitations.

TABLE III.

Probabilities of Success in 4 Sets of 8 Trials Each.

Number of Trial.

Number of Set: i \ j	1	2	3	4	5	6	7	8
1	.02	.04	.06	.08	.10	.12	.14	.16
2	.04	.06	.10	.16	.24	.34	.46	.60
3	.02	.04	.08	.14	.32	.42	.54	.68
4	.16	.20	.24	.28	.32	.40	.56	.88

Table IV.

Third Moments of Samples.

Value	Frequency * –	Frequency * +
0		147
.09375	19	6
.28125	30	29
.46975	23	19
.65625	2	2
.75	10	5
.84375	30	17
1.40625	19	22
1.5	26	17
1.96875	5	3
2.25	8	6
2.34375	1	4
2.53125	1	1
3.	14	9
3.28125	2	2
4.21875		8
4.5	1	7
6.		1
7.03125	1	2
20.25		1

* The numbers in the column headed with a minus sign are frequencies corresponding to negative values of the third moment; the numbers in the column headed by a plus sign correspond to positive values.

Table V.

Fourth Moments of Samples.

Value	Frequency
0	5
.0625	19
.08203125	26
.5	25
.76953125	59
1.	8
1.3125	15
2.5625	38
2.95703125	47
3.64453125	42
4.5	43
5.0625	9
6.64453125	2
8.	12
8.5	19
8.83203125	41
11.0625	24
11.95703125	4
12.3125	14
15.51953125	8
16.	1
19.5625	9
20.70703125	5
21.	1
21.39453125	8
22.0625	1
24.5	8
30.39453125	3
42.26953125	3
106.3125	1

HENRY D. SHELDON, JR.

Problems in the Statistical Study of Juvenile Delinquency (1)

I. — Introduction.

A statistical study of juvenile delinquency should serve three purposes: (1) to check the generality of relationships suspected from intimate study of case histories and from impressions formed in personal contact with delinquents, (2) to uncover previously unsuspected relationships which may be worth further qualitative study, and (3) to develop and evaluate techniques for achieving the first two purposes. This third aim must be added because statistical evidence, directly bearing on suspected relationships, is rarely available. Hypotheses as to factors in delinquency, as a rule, can be checked and discovered only in a round about and devious fashion.

The present study, making use of records of 6,903 male juvenile delinquencies in Cleveland, Ohio, 1928-31 inclusive, seeks to serve these three purposes.

Among the relationships suspected by various writers to be important, the following have been subjected to a statistical examination in one form or another in the present Cleveland study:

1) Other things being equal, is juvenile delinquency a function of economic status?

(1) An appointment as Post-doctorate Research Associate at the University of Wisconsin made this study possible. The writer wishes to acknowledge his indebtedness to Dr. Samuel A. Stouffer and Dr. Kimball Young of that institution for valuable suggestions and criticism. The writer is also indebted to Dr. Charles E. Gehlke of Western Reserve University and to Howard Whipple Green of the Cleveland Health Council for aid in collecting data and to Judge Harry L. Eastman and to M. S. Laird for access to the records of the Cuyahoga County Juvenile Court.

2) Other things being equal, do children of foreign-born and Negro parentage contribute more than their share of delinquents?

3) Other things being equal, are there differences in the extent of delinquency among specific nationality groups?

4) Other things being equal, is delinquency higher among boys from disorganized families and communities than from well-organized families and communities?

5) Other things being equal, is delinquency higher among boys living close to railroad tracks, industrial zones, ravines, and other "interstitial areas"? (2)

The import of the phrase, "other things being equal" becomes apparent in the following illustration:

Suppose our impression is that children of Italian immigrants are more delinquent than children of Polish immigrants. We wish to determine whether or not this difference actually exists. To do this it would be necessary to compare the proportion of delinquents in a group of children of juvenile court age of Italian parentage with the proportion of delinquents in a similar group of Polish parentage. If such a comparison shows a difference, then the further question arises as to whether this difference is due to the intrinsic Italian or Polish character of the groups compared, or it may be due to differences in economic status, family organization, or other factors. To answer this question it would be necessary to compare groups of

---

(2) These categorical questions, in the interest of clearly defining the issues, over-simplify the question of causality. In point of fact there are many conflicting points of view and much conflicting evidence on these issues. These are adequately summarized in E. H. SUTHERLAND'S *Criminology*, Chapters V-VII. Evidence on most of these questions is presented in W. HEALY and A. F. BRONNER, *Delinquents and Criminals — Their Making and Unmaking* in their analysis of extensive series of delinquents, both in Chicago and Boston. The theory of particular interest here had its origin at the University of Chicago. Areas of "deterioration" or "disorganization" within the city are postulated. These areas are low in economic status, contain a predominantly foreign-born population, possess little or no community morale, and are adjacent to railroad yards and industrial property. The relation of these areas to delinquency is described in detail in C. R. SHAW, *Delinquency Areas*, and sketched in general by R. E. PARK, "Community Organisation and Delinquency", Chapter V in R. E. PARK and E. W. BURGESS, *The City*. Personal and social disorganization as related to the immigrant and to delinquency is described by W. I. THOMAS and F. ZNANIECKI, in *The Polish Peasant*, Part IV, Section II. The rôle of railroad yards and industrial areas is emphasized by F. M. THRASHER, in *The Gang*.

Italian and Polish children roughly uniform as to these other factors. To answer it exhaustively and completely, comparison should be made in all sub-classes of one factor, other factors held constant, and then this analysis extended to the remaining factors, each in their turn. In short, a complete cross-classification of the original groups of Italian and Polish children, with respect to all factors considered as causal, is implied. If, in most of the sub-classes, there appears to be a higher proportion of delinquents among Italians, and if this difference is sufficiently large to occur infrequently by chance, (3) then the conclusion that Italians *per se* are more delinquent is justified.

The practical difficulty is that while considerable information is available for the classification of delinquents according to this ideal scheme, similar information is not available for Italian and Polish juvenile populations. In the face of this difficulty the census tract becomes a useful but imperfect device for control of factors considered as causal. As a small geographical area, the census tract limits a group of persons. It is possible to describe the characteristics of these persons in terms of delinquency and dependency rates, median rentals, percentage of homes owned, and the like.

The census tract as an instrument of this sort suffers from two major defects : (1) If used as a unit of comparison, complete homogeneity is assumed implicitly, but does not exist within the tract. (4) (2) The number of tracts is insufficient for complete cross-classification. (5)

---

(3) The determination of whether or not observed differences in proportions or rates of delinquency would happen by chance alone frequently or infrequently offers some rather special difficulties which will be given later discussion, (see p. 214).

(4) The question as to how far results are vitiated by this lack of homogeneity will be discussed later, (see p. 211-214).

(5) This would also be true of the ideal scheme outlined. In cross-classification the number of classes increases in geometric ratio. Thus if there were three factors each divided into ten groups the number of classes would be $10 \times 10 \times 10 = 1000$. This would mean that if all our cases of delinquencies were to be classified there would be on the average less than seven delinquencies to a class actually, and a large number of classes without frequencies. This effect would come into play earlier in the case of tracts, since there were only 184 of them.

This latter difficulty was met in part by the use of partial correlation as a means of controlling factors. The use of this method, however, introduces further difficulties : it requires the erection of rather elaborate logical superstructure and is valid only when data fit this superstructure. Moreover it is easy to lose sight of meanings and to misinterpret findings which become more and more involved as the number of variables increases.

In analysing the Cleveland data, the method of partial correlation was first used. The apparent findings were then examined and amplified by an intimate comparison of tracts.

II. — The data.

In the Cleveland metropolitan area there were in 1930, 252 census tracts. These are not political subdivisions but simply the smallest areas for which census summaries exist. This total included tracts for the city of Cleveland proper and the four principal suburbs : East Cleveland, Cleveland Heights, Shaker Heights, and Lakewood. It was found that with one or two exceptions (6) delinquency in the suburban cities was negligible, and since figures in some series were not available for these cities the study was confined to Cleveland proper. This left 201 tracts. Of these one was eliminated, (7) and a number were combined (8) to insure a minimum of at least 80 boys 10-17 years of age in each tract. The combinations were made on the basis of contiguity, and an effort was made to combine tracts approximately homogeneous with respect to available indexes. These combined tracts fall into two groups : tracts in the central business district of the city and tracts on the outskirts of the city. There were then 184 tracts with which to work.

Six series were selected for the partial correlation analysis. There were male delinquency, 1928-31 ; (9) home ownership,

---

(6) Notably LW–17 — an industrial area high in delinquency and percentage of foreign-born, and low in economic status.

(7) Z–1 – population 85 ; families, 23 ; no delinquents nor dependents.

(8) G–1, G–6 and G–8 ; G–2 and G–3 ; G–4 ana G–5 ; G–7 and G–9 ; H–1 and H–2 ; I–1, I–4, J–1 and J–2 ; T–8 and U–7 ; U–5 and U–6 ; W–1 and W–4 ; W–2 and W–5 ; W–7 and W–8 ; X–2 and X–3 ; X–5 and X–6.

(9) *Male Delinquency Rate.* With the aid of the Department of Sociology of Western Reserve University and the Cleveland Health Council, all cases of male delinquency for the five city areas for the years 1928–31 were

1930; (10) equivalent rental, 1930; (11) dependency, 1928; (12) unemployment, 1930-31; (13) native white of native parentage, 1930. (14)

---

tabulated. The following information was noted: address, offense number, family case number, case numbers of co-delinquents, previous offense number, offense, race, birthplace of father and mother, and age. Each offense was then allocated to a census tract, and a delinquency rate expressed as the annual average of offenses per thousand boys 10–17 years of age in 1930 was computed. This rate, when plotted against other variables, tended to produce, in some cases, fan-shaped distributions, and in others curvilinear distributions. It was found that a reasonably linear relationship was obtained by plotting the logarithms of the tract rates. A similar examination of the relationship among the other indexes indicated that several series behaved in a similar fashion; and in consequence, they were converted into logarithms for purposes of correlation analysis. (10–17 male population in H. W. Green, *Population Characteristics by Census Tracts*, Cleveland, Ohio, 1930, Table 1, pp. 73–123).

(10) *Percentage of Homes Owned*, 1930. The basic figures for this series come from the 1930 census. The number of families owning homes in which they live is expressed as a percentage of the total number of families in the given tract. A home is classified as owned "if it is owned wholly or in part by any related member of the family living in the home. A home owned by a lodger, however, is counted as rented." This percentage was used directly in all calculations. (Series in *Ibid.*, p. 60, Table 10, pp. 208–214).

(11) *Equivalent Monthly Rental*, 1930. This figure is the median of a distribution of rented dwellings combined with owned dwellings whose value has been equated to rental. The logarithms of this series were used in all calculations (Series in *Ibid.*, 64).

(12) *Dependency Rate*, 1928. Cases know to the Assocaited Charities Jewish Social Service Bureau, Mother's Pension, and Red Cross allocated to tracts were furnished by Mr. Green. A rate per 1000 families was computed for each tract. The logarithms of these rates were used in calculations. (Family population of tracts in *Ibid.*, Table 9, pp. 202–207).

(13) *Percentage Unemployment*, 1930, 1931. The number of persons unemployed in each tract was expressed as a percentage of the number of persons reporting themselves gainful workers in each tract, for the two years respectively. The two series were converted into logarithms. The correlation between them was found to be, 80. They were combined into a single index, (the average of the logartihms), and appear as such in subsequent calculations. (Two series in *Ibid.*, pp. 46–47.)

(14) *Percentage of Native White of Native Parentage*, 1930. Persons, native white of native white parentage expressed as a percentage of the total population of each area. This percentage figure appears in subsequent calculations in natural numbers. (Series in *Ibid.*, pp. 30–32,)

III. — Correlation analysis.

TABLE I. *Intercorrelation of the six selected series.* (15)

		Delinquency 1	Homes Owned 2	Rental 3	Dependency 4	Unemployment 5	Native White 6
Delinquency	1		—.70	—.58	.75	.69	—.51
Homes Owned	2	—.70		.27	—.59	—.46	.18
Rental	3	—.58	.27		—.73	—.81	.72
Dependency	4	.75	—.59	—.73		.82	—.60
Unemployment	5	.69	—.46	—.81	.82		—.71
Native White	6	—.51	.18	.72	—.60	—.71	

An examination of Table I indicates that all five independent variables show a moderately high relation ship with delinquency, but it will be noted that with the exception of home ownership about the same degree of relationship exists among themselves. This situation suggests that there is a certain element of duplication in these series. Since, as we have noted, it is desirable to keep the number of variables at a minumum, it was of interest to discover whether any of these variables might be eliminated.

To this end a multiple regression equation predicting delinquency from the five independent variables was computed :

$$X_1 = 1.3343 + (-.934235)(X_2) + (-.228746)(X_3) + (.219.826)(X_4) + (.199194)(X_5) + (-.366082)(X_6).$$

While none of the variables appear precisely negligible, it seemed questionable whether the control gained by the inclusion of all of them compensated for the complexity introduced into the analysis. The multiple correlation between delinquency and the other five

(15) It should be noted that these coefficients obtain only for the particular areas used. In general, it appears that if the size of areas is increased (and the number of areas, in consequence, decreased) observed correlations tend to increase. The problem here, analogous to that in agricultural experiments, is the determination of an optimum area-size which insures maximum stability of indexes, and at the same time, minimum heterogeneity. Apparently this problem can be solved only in an empirical and common-sense fashion. Since in agricultural experimentation a much greater degree of control is possible, more confidence can be placed in an empirical solution. In the present study the smallest available units of area have been used. With exceptions, the indexes seem fairly stable ; and heterogeneity, although by no means eliminated, has been reduced in this respect to the lowest level possible in terms of the data.

variables was found to be .84, which, incidentally, indicates a rather high degree of control for this type of study. But it was found that the multiple correlation of delinquency with home ownership and rental was .80. If the objective, then, is merely to obtain a reasonably close fitting " prediction " (16) equation for delinquency, this can be done about as readily from a knowledge of rental and home ownership as from a knowledge of all five variables. Close " prediction " is not precisely our objective, however; we are concerned rather with the importance of each of these factors in delinquency, " other things being equal. " Table II presents evidence on this point.

TABLE II. *Partial correlation coefficients: (X_1) delinquency and (X_2) home ownership, $(X_3 X_5)$ rental and unemployment, (X_4) dependency, and (X_6) native white.*

$r12$	—.70	$R1.35$	.68	$r14$	.75	$r16$	—.51
$r12.3$	—.69	$^*R1.35.2$	.61	$r14.2$	.59	$r16.2$	—.56
$r12.34$	—.52	$^*R1.35.24$	.29	$r14.23$	.31	$r16.23$	—.26
$r12.345$	—.51	$^*R1.35.246$	.16	$r14.235$	.22	$r16.234$	—.21
$r12.3456$	—.53			$r14.2356$	.21	$r16.2345$	—.18

From Table II some interesting observations may be made:

1) The factor of home ownership is quite closely associated with juvenile delinquency. Even with four other factors held constant the partial correlation stands at .53.

2) As compared with what previous less controlled studies of delinquency have suggested, the factor of nativity has a surprisingly low association, with the other factors held constant. The final correlation is —.18, which, however, seems to be significant since a

---

(16) " Prediction " here is used in the sense of accounting for a maximum amount of delinquency in terms of other variables at the given time of the study and has no implications for predicting delinquency in the future.

(*) Representing the multiple correlation between juvenile delinquency and rental and unemployment after holding constant the variable or variables indicated in the second set of subscripts. Hereafter in the text a correlation of this type is called a combined partial correlation. The formula may be written:

$$R\ 1.35.246 = \sqrt{1 - (1 - r^2\ 13.246)\ (1 - r^2\ 15.2346)}\ .$$

For a discussion of this new measure see SAMUEL A. STOUFFER, *A Coefficient of combined Partial Correlation* « Journal of the American Statistical Association », March, 1934.

negative coefficient of this magnitude would would occur by chance less than once in 50 times.

3) Rental, unemployment, and dependency were originally used as indexes of economic status. Their "combined partial" correlation with juvenile delinquency, with other factors held constant, $R1.345.26$, was found to be .40 (not included in Table II). These factors taken together are clearly important. Dependency, however, seemed to differ markedly from the other two indexes in its relation to juvenile delinquency. The partial correlation between delinquency and dependency, all other factors constant, is +.21, which, while rather low, could happen (with a positive sign) less than once in 100 times by chance. It is evident, then, that even if areas had the same economic conditions as measured by rental and unemployment and also had the same home ownership and nativity, those districts with more dependency would have more delinquency. What is "dependency, other indexes of economic status constant" really measuring? This question will be discussed subsequently. After taking out dependency, the relation between delinquency and rental and unemployment is less marked. With all other factors constant, the "combined partial" correlation $R1.35.246$ drops to .16, which is low, though clearly on the verge of significance.

It would be expected, then, in considering individual areas in Cleveland that, *other factors held constant*, areas high in home ownership would be low in delinquency, that areas high in dependency and low in other economic indexes would be high in delinquency, and that large foreign-born and Negro elements in the population of a given area might, on occasion, be associated with a high delinquency rate.

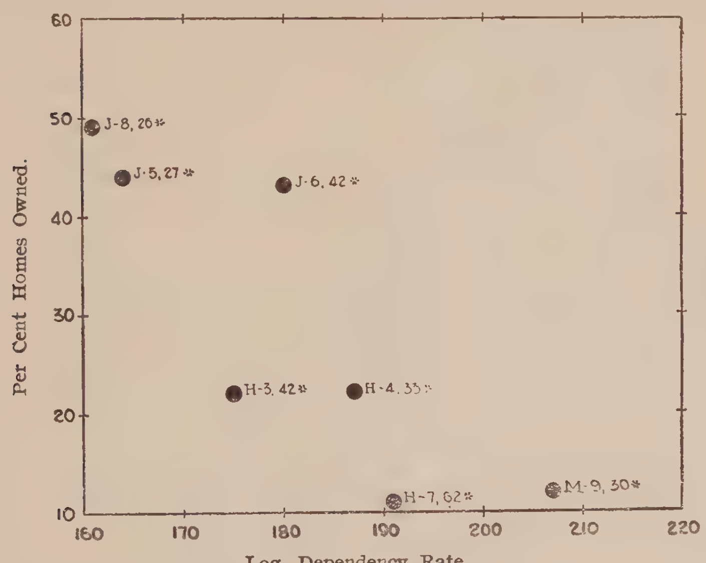
IV. — Selection and comparison of divergent tracts.

The preceding partial correlation analysis provides a general pattern of delinquency expectation, but it has by no means accounted for all delinquency, nor has it defined causal factors except in terms of indexes of somewhat ambiguous reference. A detailed comparison of tracts limits this element of ambiguity and permits the examination of hypotheses which cannot be framed in quantitative terms. Such comparison in itself, however, is often meaningless unless its relationship to the general pattern can be specified. Using this general pattern set up by the correlation analysis as a frame of reference, a detailed comparison of certain selected tracts was made.

The tracts, arrayed with respect to percentage of native whites were divided into quartiles. Each of these groups was divided, in turn, into fifteen sub-classes defined in terms of an economic index, (17) thus creating some sixty categories in which economic status and nativity were held relatively constant. Categories in which more than four tracts fell were selected for further analysis. The tracts in each of the categories were plotted on scatter diagrams indicating the relationship between home ownership and dependency. The delinquency rate for each tract plotted was then written in. If the tracts conformed to the general pattern, those low in home ownership would be low in dependency and *vice versa*, and those low in home ownership and high in dependency would have high delinquency rates, while those low in dependency and high in home onership would have low delinquency rates. Tracts which failed to conform to this pattern were selected for further scrutiny. Figure I illustrates this procedure :

FIGURE I.

Distribution of Tracts with Respect to Dependency and Home Ownership.



* Delinquencies per 1,000 boys.

(17) The sum of the deviations of the rental and dependency series from their respective means deflated by their respective standard deviations (*i. e.*, a " standard score ").

In Figure I a group of tracts are presented which are relatively uniform with respect to nativity and economic status. [18] As far as dependency and home ownership are concerned, these tracts conform to the general pattern. There are, however, a number of discrepancies in the delinquency rates. *H*-7 and *M*-9 are roughly similar with respect to dependency and home ownership, but there is considerable difference in delinquency ; *J*-8, *J*-5, and *M*-9 have about the same delinquency rates, but *M*-9 is much higher in dependency and lower in home ownership. These divergent tracts, then, are indicated for further study.

This method of analysis is valuable in that it permits the isolation of tracts which are similar in respect to several factors but differ markedly in respect to one factor ; in short, it represents about the closest approach possible to a "controlled experiment" in this type of data.

A number of tracts were selected by this procedure. The writer examined them personally, and interviewed social workers, probation officers, and other persons thought to have an intimate knowledge of the areas under consideration in an attempt to explain deviations and discrepancies. While the results of this investigation were not as productive as had been anticipated, issues of considerable interest were raised.

One of the first additional factors which suggested itself as a means of explanation of the failure of this group of tracts to conform to the general observed pattern was that of nationality. While it was found in those comparisons which involved tracts predominantly Negro and predominantly white that the Negro tracts had a higher delinquency rate than expected, this was not true in all cases. Tracts involving other nationality groups showed no consistent differences. Among tracts predominantly Polish, for example, some were found with higher and other e with lower delinquency rates than would be expected from a knowledge of the five variables used in the correlation study. Consideration was also given tracts of mixed nationality composition on the assumption that delinquency might be augmented as the result of the conflicting *mores* of the several groups. Here again the results were inconclusive. At this level of evidence, then, it would seem that with the possible exception of Negroes, specific nativity in itself is not a significant factor in

(18) Percentage of native white : 13-24 ; economic index —151 to —200.

delinquency. It must be recognized that this finding is tentative and suggestive rather than conclusive.

The hypothesis that the "interstitial area" (defined in terms of railroad property) tends to augment delinquency was then examined as a means of explaining the non-conformity of tracts. As in the case of nationality, this factor did not seem to explain differences in delinquency, at least in this frame of reference. Comparable tracts, seemingly equally "interstitial" in respect to railroads, had delinquency rates both higher and lower than expected. Moreover, among comparable tracts both "interstitial" and "non-interstitial," several "non-interstitial" tracts appeared with higher delinquency rates. (19)

A comparison of tracts *Q*-5 and *M*-3 led to the consideration of a further hypothesis as to factors in delinquency rates. From the type of analysis illustrated in Figure I attention was called to discrepancies in these two tracts. They are uniform in respect to rental, percentage of native white, and unemployment. *Q*-5, however, has a much lower dependency rate and a much higher percentage of home ownership; moreover, 69 per cent of the population of *Q*-5 is Italian, while 81 per cent of the population of *M*-3 is Negro; consequently more delinquency would be expected in *M*-3, but actually the rates are approximately the same. A probation officer whose district includes *Q*-5 suggested that the rate in *Q*-5 was higher than expected because there were a few large families in the area, the boys from which were habitually delinquent.

---

(19) Considering the city as a whole, it is generally true that tracts adjacent to railroad and industrial property have the highest delinquency rates; but these are also tracts of low rental, low home ownership, low percentage of native white, and of high dependency and unemployment. Thus if the "interstitial area" is to be attributed independent efficiency as a factor in delinquency, it must be defined in terms which differentiate "interstitial" tracts from tracts of low economic status in general. A definition in terms of railroad property is undoubtedly too narrow but does afford one of the few clear-cut bases of differentiation. A satisfactory definition would have to grow out of an intimate knowledge of where boys actually gather and their migratory habits. The lake front, for example, with its docks, junk piles, and other industrial properties, seems an "interstitial area" in THRASHER'S sense. The fact that boys from many distant parts of the city come to the lake front in summer vitiates a statistical picture of delinquency in terms of continuity of the lake of tracts in which they reside. For a discussion of the concept to "interstitial areas" see THRASHER, *The Gang*.

In order to examine this hypothesis some thirty tracts, with a sufficient number of delinquencies to warrant some confidence in results, and with high and low delinquency rates, were analysed with respect to number of families from which delinquents came and number of offenses committed by the same boy during the period under consideration.

This analysis indicated, for the tracts examined, at least, that recidivism appeared to be a relatively constant element in offenses. It also indicated that while a small proportion of families contributed a large proportion of offenses (*e. g.*, in *Q*-5,30 per cent of all families contributing offenses contributed 60 per cent of all offenses) this was true of most of the tracts examined. [20]

While the analysis of these thirty tracts failed to produce adequate evidence that a small number of families contributed a large proportion of offenses in tracts in which the delinquency rate was higher than expected, it did show rather clearly that the number of families contributing delinquents seemed consistently to be a small proportion of the total number of families in the tract. These families contributing delinquents, then, could differ quite markedly from the rest of the families in the tract and still not appreciably affect other indexes for the tract. At the same time, they would have a marked effect on the delinquency rate and thus might throw this rate out of line with the other indexes. This effect would obviously, marked, be most when, in addition to differing from other families, the families contributing delinquents were, on the average, larger than other families. Tracts in which this situation obtains are cited in the following section, which deals specifically with the problem of heterogeneity within tracts.

V. — The problem of heterogeneity.

It was noted in the introduction that the use of the census tract to define groups implied homogeneity within tracts, and that complete homogeneity did not exist. That is if, one is ascribing

(20) It will be noted that the size of families contributing offenses, which is the crucial factor in an analysis of this sort, has been ignored since it was unavailable. Average family size for tracts, unrelated to recidivism or to specific families contributing delinquencies, shows little correlation with the delinquency rate ($r = .12$). This correlation in terms of boys, rather than persons, per family is similar ($r = .13$).

delinquency to economic status, to a given nationality group, or to the presence of apartment houses in a given area, it is tacitly assumed that the juvenile population lives on the average in homes of the same rental as the average for the tract, that this juvenile population lives in apartment houses to about the same degree as the rest of the inhabitants, and that the percentage of a given nationality is closely reflected in the juvenile population.

This assumption has been subjected to sharp *a priori* questioning: If 50 per cent of the families in an area own homes, does it follow that 50 per cent of the boys live in owned homes? If, in an area, 50 per cent of the inhabitants are Greeks, are boys of Greek parentage in the same proportion, or do the 50 per cent Greeks represent a community of single males? On what basis can it be assumed, that because the population of an area is 75 per cent Polish and the dependency rate is high, that in the area Poles are essentially a dependent group?

The foregoing analysis of offenses within tracts with respect to large families and recidivism indicated that there is empirical basis for questions of this sort. Table III presents a group of tracts which give further basis for questioning and specific illustration of the type of heterogeneity previously discussed.

TABLE III. — *Heterogeneous tracts (selected).*

Tract	Delinquency Rate	Dependency Rate	Percentage Homes Owned	Equivalent Rental	Percentage Unemployment 1930	1931	Percentage Native White
G-1-6-8	46	102	2	42	16	25	47
M-1	48	30	5	57	7	13	52
L-8	25	28	9	55	8	12	63
R-9	19	8	13	63	3	6	61
S-5	19	2	28	74	3	12	37
R-7	12	6	22	80	3	8	52

The first four tracts presented in Table III have the following characteristics in common: they have the lowest number of boys per family in the city; they are all hotel, apartment house, or rooming house areas, and on scatter diagrams of delinquency with percentage of native white and with rental they stand out as possessing a too high delinquency rate.

In the case of *G*-1, 6, 8, and *M*-1 the explanation of this rate

is clear: Most of their population is a hotel and apartment house population, relatively prosperous and predominantly native white but in the main childless. *M*-1 is adjacent to a Negro area and *G*-1, 6, 8 to foreign-born areas; the population of these adjacent areas has infiltrated into the tracts under consideration, introducing a minority which differs radically from the populations of the tracts as a whole. This element contributes the delinquencies. In *M*-1 21 out of 25 offenses were contributed by Negro or foreign-born families and in *G*-1,6,8, 13 out of 15 offenses were contributed by the same element. (21)

In contrast, *L*-8 and *R*-9, while generally similar to the preceding tracts, are not adjacent to Negro or foreign-born areas, nor do these elements appear among their delinquencies. In consequence, if a heterogeneous element contributing delinquencies exists, there are no data available with which to specify such an element.

In tracts *S*-5 and *R*-7 the delinquency rate is high with respect to percentage of native white. *S*-5, a high-class residential area, contains apartment houses whose rentals are among the highest in the city, but it also has a fringe of two-family houses. *R*-7 contains a central core of fine residences and a periphery of apartment and two-family houses of much lower rental. The high-class residential areas contribute the rental figure and the remaining sections the delinquents. (22)

In summary then, in has been noted that the use of census tracts as units assumed internal homogeneity, that this assumption might be readily questioned, that an examination of some thirty tracts in respect to proportion of families contributing delinquents provided empirical grounds for such question, and that certain tracts failed to behave as expected as a result of heterogeneity. The seriousness of this difficulty depends on the investigator's objective. If his aim is a precise and exhaustive mathematical analysis of fac-

---

(21) Emphasis should be placed, not on nativity alone, as the explanatory factor, but rather on its probable association with low economic status, low home ownership, and the like.

(22) In the case of *S*–5 the two-family element can again be defined in terms of nationality. The section in question is inhabited largely by Hungarians. Of the 14 offenses in the tract 12 were contributed by families of foreign origin, and 7 of these 12 by families of Hungarian origin. In the case of *R*–7, however, 5 of the 6 delinquencies came from native white families.

tors in delinquency, heterogeneity within tracts would seem to prove a serious, if not insurmountable, obstacle. If, on the contrary, this aim is a rough evaluation of alleged causal factors in terms of mass phenomena, and the development of a frame of reference for detailed case and community studies, heterogeneity is by no means so serious a difficulty. The correlations observed among the various series, although they could scarcely be invested with ultimate significance, were sufficiently high to indicate that, in a measure, in areas where a large number of homes were owned, the number of delinquencies was small, and in areas where the equivalent rental was the low, number of delinquencies was large. It seems probable also that there exists a good deal of overlapping, that is, that a fair share of non-delinquents live in owned homes and that delinquents come from homes whose rental is no higher than that of other homes in the tract. (23) If attention is focused on specific tracts, heterogeneity becomes a greater hazard, but the detaile consideration of these tracts usually leads to the discovery of heterogeneity, and thus it can be taken into account. Moreover, its discovery in the specific instance often opens the way to refinement of indexes and thus makes possible the reduction of heterogeneity in the general study. In short, heterogeneity very probably prevents the reduction of factors in delinquency to a mathematical formula. It does not, however, prevent productive study of these factors by the use of census tracts.

VI. — Problem of significant differences in delinquency rates.

It was indicated in the introduction that an ideal scheme for testing hypotheses as to factors in delinquency involved some demonstration that observed differences were sufficiently great to happen infrequently by chance alone. The fact that sub-classification reduces the number of frequencies in any category makes some demonstration of this sort imperative. This is particularly true of offenses ; the reader has probably noted in the discussion of heterogeneity that, in the tracts selected, the highest number of offenses

---

(23) Another view might be taken of this matter : namely, that home ownership or dependency, in addition to indicating the immediate home environment of boys of juvenile court age are also imperfect indexes of a local cultural or psychological milieu, which is the crucial factor in delinquency. In this case the question of heterogeneity becomes unimportant except in cases where the tract encompasses several distinct types of cummunities.

was 25 and asked the question whether the alleged "highness" of the delinquency rates had any meaning, or was merely a chance "highness."

Unfortunately, such a demonstration seems impossible in terms of present data and present techniques. All error formulae are based on the assumption of independent events. This implies for delinquency that in a given category each offense occurs independently of every other offense. Obviously this condition is not met, since a given offense is related to and conditioned by previous offenses, other offenders in the family, and associates in group offenses. This difficulty could be overcome only partially by making comparisons in terms of single offenders. In the correlation analysis, however, enough variables are held constant to give one some confidence in considering the residuals from the multiple regression plane sufficiently independent to provide a basis for estimating chance errors of regression coefficients. Likewise the computation of the chance error of the partial correlation coefficients seems to be defensible.

VII. — Consideration of indexes.

In this section are considered certain findings relevant to an evaluation of some of the series used in the partial correlation study.

Delinquency. The analysis of selected tracts with respect to family size and recidivism certainly raised the question as to what constitutes the most adequate index of delinquency. A rate of families contributing delinquents per 1,000 families proved to be impractical. [24] The logical rate would seem to be, of course, that of delinquents per 1,000 boys (rather than delinquencies per 1,000 boys) ; on the other hand, if one is interested in the total problem of delinquency, recidivists are as important as first offenders.

The examination of these selected tracts, however, seemed to indicate that it made little difference whether rates were computed in terms of boys or offenses. That is, in tracts where the number of offenses was fairly large a relatively constant ratïo seemed to obtain between boys and offenses. This stability, from another angle,

(24) It was found that this rate was determined more by the number of one-and two-member families than by the number of families contributing delinquents. If the one-and two-person families were ruled out, there still remained the problem of making adjustment for the differential size of the remaining families.

was indicated further by a consideration of the relationships obtaining among rates computed from total offenses, offenses against property, and group offenses. The correlation coefficients among these rates were found to be .90 or higher. The correlation between rates based on the 1928-29 and 1930-31 delinquencies respectively was found to be .86, which gives added confidence as to the stability of the delinquency series.

Home Ownership. This series was defined in terms of the percentage of families owning their own homes in a given area. If, however, an area contains only two-family houses the maximum percentage of ownership would be 50 ; if an area contained only four-family apartments this maximum would drop to 25 per cent. It might reasonably be supposed that the renters in two-family areas are essentially of the same character as the owner in these areas and thus the original series under-estimates home ownership. In view of this supposition another index was obtained, home ownership per 100 dwellings, (25) Since, is this index, a dwelling is defined as a building, 100 per cent ownership is possible in all areas, and the difficulty in two-family areas in corrected. (26) This second index was correlated with other variables used in the partial correlation analysis with the following results :

	Delinquency Rate	Percentage Homes Owned	Equivalent Rental	Dependency Rate	Percentage Unemployment	Percentage Native White
Percentage Homes Owned	—.70		.27	—.59	—.46	.18
Homes Owned per 100 Dwellings .	—.68	.85	.28	—.62	—.50	.10

It would seem, then, that nothing was to be gained statistically by defining home ownership in different terms.

Percentage of Native White. It will be noted that in the use of this series the assumption is that the percentage of native white of native parentage tends to reduce delinquency, while percentage of foreign or Negro parentage tends to increase delinquency. Among

(25) H. W. Green, *Real Property Inventory of the Cleveland Metropolitan Districts*, p. 7.

(26) Also in other areas. However, in apartment-house and slum areas it is the apartment or slum character of the area rather than the proportion of resident owners which is important.

the foreign-born group, however, are included native white of German and British parentage This group, culturally nearest akin to the native white group, and probably most easily assimilated, might well be regarded as an element tending to reduce rather than raise delinquency. In view of this theory the percentage was computed for each tract. This series was then correlated with the original series with the following results:

	Delinquency Rate	Percentage Homes Owned	Equivalent Rental	Dependency Rate	Percentage Unemployment	Percentage Native White
Percentage Native White of Native Parentage . . .	—.51	.18	.72	—.60	—.71	
Percenage Nattive, British, and German Parentage . .	—.57	.24	.70	—.63	—.70	.97

Here again a refinement of the native white series on adequate *a priori* grounds failed to produce a substantial change in the relationship of this index to the other indexes considered in the study.

Dependency. Another dependency series for the year 1931 was obtained after the correlation study had been completed; therefore it was not used in the original computations. It has, however, been correlated with the original series:

	Delinquency Rate	Percentage Homes Owned	Equivalent Rental	Dependency Rate 1928	Percentage Unemployment	Percentage Native White
Dependency 1928	.75	—.59	—.73		.82	—.60
Dependency 1931	.77	—.56	—.77	.90	.86	—.64

These correlations seem to indicate a rather consistent behavior on the part of dependency both in a period of prosperity and in one of depression and thus permit added confidence in the dependency.

A comparison of predominantly Jewish tracts with non-Jewish tracts of similar economic status in terms of rental and unemployment revealed disproportionately high rates of dependency in the Jewish tracts. Informants and available figures failed to give adequate explanation of this discrepancy, but the general impression

seemed to be that the differences observed reflected differences between the administrative and recording policies of the Jewish Social Service Bureau and the Associated Charities rather than actual differences in dependency.

Equivalent Rental. A comparison of tracts predominantly Negro in population with non-Negro tracts of similar character seemed to indicate that Negroes paid higher rental than whites. The evidence for this supposition lay in the fact that while rental in the Negro tracts was generally higher than the rental in the white tracts, at the same time, contrary to expectation, dependency and unemployment were also higher in the Negro tracts. This inflation of rental was also found to be true of certain rooming-house and apartment areas.

This examination of series points to two conclusions: (1) that in a general description in terms of the correlation of indexes, these indexes show a surprising stability, if not inertia; (2) that, while these indexes appear stable when considered *in toto*, when attention is focused on a given index of a specific area, it is necessary to determine precisely the character of that index before using it as a basis for generalization.

Working Concepts vs.

Theoretical Concepts. One of the chief difficulties in a study of this sort is that of bridging the gap between theoretical concepts and the index or working concepts necessary for statistical purposes. The three factors considered in the correlation analysis have been economic status, nativity, and community and family disorganization. The relationship between economic status and rental, dependency, and unemployment is sufficiently clear to obviate further discussion. Likewise, nativity, while it has implications for disorganization, is a theoretical issue in its own right.

While adequate definitions of family and community disorganization as theoretical concepts are possible only in terms of elaborate and extended dialectic, their general sense may be indicated briefly. A *disorganized family* may be considered as a family which has not made adequate adjustment to its social environment. In a *disorganized area* the behavior of individuals is not regulated by social controls proceeding from a single integrated pattern of mores, attitudes, and values. Here own standardes not in harmony with that of the larger community. That is, community disorganization implies either the absence of such a pattern, the existence of competing

and conflicting patterns, or a single pattern which does not conform to that of the larger community.

Obviously these definitions do not lend themselves readily to statistical analysis. It becomes necessary then to turn to indexes which appear to reflect the central psychological and sociological core of disorganization. That is, for a working concept one is reduced to symptoms of disorganization which can be expressed quantitatively. In terms of a working concept a *disorganized family* is characterized by dependency, parental conflict, desertion, drunkenness, immorality and the like. A *disorganized area* is an area of high dependency, unemployment, density, percentage of foreign-born and Negro, and mobility ; of low economic status and home ownership, of decreasing population ; and an area usually adjacent to railroad, industrial, or business property. (27)

For the purposes of this study two series were selected as a basis for a working concept of disorganization : home ownership and dependency. Home ownership was selected on the ground that families able and desiring to own homes were better able to manage their children, and on the ground that communities in which homes were owned possessed a more stable community morale. Dependency was thought to have similar implications in addition to its obvious relation to economic status. (28)

VIII. — Conclusions as to delinquency.

1. The three series (rental, unemployment, and dependency) taken as indicative of economic status considered separately, other factors held constant, showed no very close relationship to delinquency. This result was to be expected, since they were so highly correlated among themselves. When these factors, however, were combined in a " combined partial " correlation, with home ownership

---

(27) This list, of course, represents a catalogue of traits which have been attributed to the disorganized area by various writers. Actually it would be extremely difficult to discover any great number of tracts which fitted the definition on all counts.

(28) Other series might have been used. An adequate index of broken homes, if available, might have been valuable, although it appears that dependency covers much the same ground. Other indexes such as population increase or decrease and density were discarded since their bearing on disorganization seemed less direct than the above series, and scatter diagrams showed them to behave in approximately the same fashion as the series used.

and percentage of native white held constant, the coefficient became .40. This would indicate economic status as a significant and important factor in delinquency. (29)

2. The series used as indexes of family and social disorganization in both cases showed a significant relationship with delinquency. The partial correlation between delinquency and home ownership was — .53 and that between dependency and delinquency was .21. This latter coefficient, while not so strikingly large as the former, is significant. A " combined partial " correlation between delinquency and home ownership, other factors held constant, would seem to indicate the total effect of disorganization on delinquency, that is, the effect of home ownership plus that part of dependency which appears to be non-economic. The value of this coefficient was found to be .66. (30) Thus it would appear on the basis of this correlation analysis that social disorganization is the most important single factor in delinquency. This finding more than any other partakes of the nature of a genuine positive contribution.

3. The partial correlation of percentage of native white with delinquency was found to be —.18, which is significant but lower than might be expected from the evidence on this question which has grown out of case studies.

4. A series of exploratory forays made by the comparison of tracts which did not conform to the general pattern of expectation failed to produce any evidence that such factors as nationality, interstitial areas, size of family, or recidivism were efficient. This analysis, however, did not demonstrate that these factors were irrelevant. In short, the evidence on these issues was inconclusive.

IX. — Conclusions as to methodology.

1. Correlation analysis is an effective tool for constructing a general pattern of delinquency expectation. Such a frame of reference is almost imperative if the results of detailed case and historical studies are to be genuinely significant and meaningful. A general

---

(29) Dependency seems to involve an element not strictly economic. By holding home ownership constant in obtaining the correlation of .40 it is thought that allowance is made for the non-economic element in dependency. (See p. 207).

(30) Again by holding economic factors constant it was thought that that part of dependency with economic implications was held constant.

description of this sort serves as a point of reference both for selecting problems for detailed study and checking and evaluating the results of such study.

2. Correlation analysis has definite limitations: (*a*) Since it is based on units of area, general and final significance cannot be attributed to coefficients computed; (*b*) since it describes the situation "on the average," significant detail is always obscured; (*c*) since it is necessary to isolate the independent causal efficiencies of factors considered, partial correlation is essential. Successful and clear-cut interpretation of partial correlation becomes increasingly difficult as the number of variables is increased.

3. Within these limitations this study has indicated a reasonable basis for a considerable confidence in the results of such analysis. The multiple correlation between delinquency and the independent variables considered was found to be .84. As noted, this figure indicates a rather high degree of control for data of the type used. The correlation between the delinquency rates for 1928-29 and 1930-31 was found to be .86 and that between the dependency rates for 1928 and 1931 was found to be .90. These coefficients indicate a suprising stability in the two indexes which on *a priori* grounds seem susceptible to the greatest fluctuations. Stability is also reflected in the high relationships observed among delinquency rates whether based on total offenses, offences against property, or group offenses, and in the similarity of behavior, with respect to all variables, of two somewhat different indexes of home ownership and of a similar pair of indexes of nativity.

4. The method illustrated in Figure 1 is valuable in selecting divergent tracts which in turn indicate problems for intensive study, heterogeneity, and defective indexes.

5. The presence of heterogeneity and defective indexes seems inevitable. This situation imposes a serious limitation on any complete large scale statistical analysis of urban communities. It also means that in a detailed comparison of tracts account must be taken of these factors before the comparison can be used as evidence bearing on any given hypothesis.

6. In the comparison of groups defined in term of tracts or otherwise, it is extremely important to specify the rôle of chance in observed differences. With respect to delinquency such a procedure requires either special techniques or special types of data, neither of which are at the present time available.

7. The chief methodological contribution of this study is the development of a method by which (*a*) a general pattern of delinquency for an entire city is described ; (*b*) tracts which do not conform to this pattern are isolated ; (*c*) heterogeneity and defective indexes may be identified and hypotheses may be given preliminary testing. The ground is then cleared for intensive historical study of crucial areas and case study of delinquents in these areas. The results of such a study would be particularly valuable, since it would be a study of a problem significant in terms of the general pattern, and the results could be oriented toward that pattern. Moreover, a considerable increase in knowledge might be expected from the interaction of these two methods. The intimate study would point the way to improvement and refinement of indexes which would make possible greater refinement in a general statistical analysis which in turn would specify more clearly problems for further intimate study.

PROF. CORRADO GINI, *Direttore responsabile.*

Roma — « L'Universale » Tipografia Poliglotta.